Dottorato di Ricerca XXXIII ciclo – Università degli Studi di Genova

PhD candidate:

Matteo Cecchetto

Title:

Structure and temporal dynamics of marine biodiversity in the Terra Nova Bay area (Ross Sea, Antarctica) analysed through DNA barcoding and metabarcoding: A standardized and systematic approach for the construction of a “DNA Barcode” library.

Supervisor:

Prof. Stefano Schiaparelli

Aims and structure of the thesis:

The firs objective of this thesis consisted in the quantification of the reference library completeness for the Cytochrome c oxidase subunit I (COI) barcode of metazoan species occurring in the Ross Sea MPA and in identifying which taxonomic groups in the last decades were investigated by using “DNA barcode” techniques performed in the Southern Ocean. In order to do that, all the available sequence data on major public repositories were gathered and analyzed. A collection of COI sequences amplified from specimen of the Italian National Antarctic Museum (MNA) was produced and aggregated to the latter to assemble the first global Ross Sea reference library of metazoan COI barcoding sequences.

The potentialities of “DNA metabarcoding” techniques applied to the analysis of Antarctic biological communities were also evaluated in Terra Nova Bay (TNB) by focusing on the development of macrozoobenthic pioneering communities colonizing artificial structures over a period of three years and on nanoplankton temporal short-term dynamics in two consecutive years. These studies were conducted using innovative sampling methodologies and experiments and differed not only for the investigated community, but also for the sampling frequency, timing and study purposes. Due to the peculiarities of the three analyses, these will be detailed in three separate chapters.

Each chapter is composed by its specific “Introduction” to the concepts discussed in that study, followed by a section on “Materials and Methods” and a joint section with the “Results and Discussion”. Considering that the different chapters, notwithstanding the fact that they pertain to the same, general topic of research, greatly differs in the overall design, as mentioned earlier, no general Discussion for the entire thesis was produced. Instead, an Introduction, discussing the most general topics covered by the entire thesis, is presented at first (*i.e.* the “Thesis Introduction”), whereas the thesis’ Conclusions, considering all the outcomes reported in the different chapters, will be presented at the end of the whole thesis (“Thesis Conclusions”).

Thesis Introduction

Antarctic marine biodiversity can be considered as the result of a combination of very peculiar environmental conditions (Peck, 2018). The origin of these environmental conditions dates back to the latest part of the Cenozoic era, during which the numerous glacial-interglacial cycles, together with the isolation of the continent from all the other land masses, influenced the Antarctic ecosystems at a degree that has no parallel on Earth (Thatje et al., 2005). More than 8,000 species have been described for the Southern Ocean (De Broyer et al., 2014) and molecular techniques are identifying new, cryptic species, increasing the knowledge on the true diversity of this region (Convey and Peck, 2019). However, the low accessibility of the entire continent has led to uneven and sporadic sampling, leaving areas still not properly investigated and meaning that the true diversity of Antarctic marine communities might still be unknown (Convey and Peck, 2019). Many oceanographic campaigns have been led in the last decades in order to increase the knowledge on Antarctic communities, but still, the lack of exhaustive information on the recurrence and distribution of species in the Southern Ocean represents a significant impediment to a proper understanding of the biological component of this ecosystem.

In this context, we are aware that major changes are affecting oceans' functioning, but no coordinated effort has been yet undertaken in order to understand how and to what extent the effects of these environmental changes will take place (Bindoff et al., 2019). A recent literature review focusing on the topic "Antarctic biotas in transition as a response to environmental change" undertaken by 26 experts (Gutt et al., 2020) revealed that 67% of the relative papers were published in the last decade (2010-2020), indicating that the majority of the available information on Antarctic communities, at least in respect to environmental change, has been gathered only recently.

In the Southern Ocean, major drivers such as the increase of temperatures, ocean acidification and altered sea ice dynamics are expected to be the most important factors influencing the future biological communities (Convey and Peck, 2019). Here, therefore, there is an increasing need of long-

term monitoring programs, especially in a multidisciplinary setting (Convey and Peck, 2019), where fine-scale approaches would be useful to track changes at high resolution. However, long-term biodiversity monitoring programs have been rarely performed in Antarctica, especially regarding microbial or benthic communities (Jones et al., 2020), with only some, out of the ordinary monitoring activities protracted for decades (*e.g.* Dayton et al., 2016).

A partial explanation for this deficiency in Antarctic bio-monitoring research must be identified in the logistic constraints, which are often driven by financial shortcomings and particularly relevant when sampling takes place in remote areas (Lacoursière Roussel et al., 2018). In fact, polar areas are characterized by uncomfortable environmental settings, with temperatures exceeding the freezing point exclusively during the summer months, sea ice cover changing abruptly in a short period of time and harsh weather conditions, inevitably obstructing sampling activities. On top of this there is also the higher cost of maintaining personnel in these remote areas. Thus, when operating in Antarctica, the fulfilment of one of the most important requirements of a sound monitoring program, *i.e.* a high sampling frequency, is generally difficult to be achieved (Proença et al., 2017).

Considering these circumstances, traditional methods relying on morphological identification have failed to provide an appropriate solution to these issues (Chain et al., 2016; Gast et al., 2006; Zhang et al., 2018). These methods, in fact, require a lengthy period of sample processing time and, in consequence, are generally used on a local-scale, thus leading to higher costs and a magnification of all the above issues (Baird and Hajibabaei, 2012). Moreover, they are also affected by low precision and reproducibility (Baird and Hajibabaei, 2012). One solution proposed to overcome this problem relies on the use of High Throughput Sequencing (HTS), which gained more attention in the last decade due to its high reproducibility, short period of processing time and steadily decreasing costs of the analyses (*e.g.* Valentini et al., 2016; West et al., 2020). The application of HTS technologies to biodiversity research, thanks to the aforementioned advantages, has been extensively referred to as

"DNA metabarcoding" (Taberlet et al., 2012) and gained increasing interest in the last decade (Taberlet et al., 2018).

These technologies have been widely adopted for a variety of different research purposes, from single species detection (*e.g.* Ficetola et al., 2008) to the evaluation of entire living communities (*e.g.* Wangensteen et al., 2018) and also to reconstruct past communities through the amplification and sequencing of sedimentary DNA (Capo et al., 2021; Edwards, 2020). "DNA metabarcoding" has proved to be extremely useful in bio-monitoring research, for both the evaluation of taxonomic composition changes (*e.g.* Martin-Platero et al., 2018) to the detection of invasive species in both the present (Westfall et al., 2020) and the past (Ficetola et al., 2018). However, these technologies present some disadvantages that still force us to limit the potentialities of such methodologies (Taberlet et al., 2018). The quantification of abundances from amplicon sequencing data is hampered by different issues mainly related to PCR amplification biases (Piñol et al., 2018), for example, but such biases may be resolved by the adoption of different methodologies not relying on PCR amplification (Armbrecht, 2020).

Nonetheless, one of the most relevant issues regarding "DNA metabarcoding" is the low reference library completeness (Taberlet et al., 2018). In fact, the automatic taxonomic assignment of the sequences retrieved using these technologies is based on the use of different databases which are composed of reference sequences, usually obtained through "DNA barcoding" techniques and deposited in continuously updated online databases, that have been extracted from specimens identified by taxonomic experts. Many different long-term projects have been conducted in order to facilitate and boost the accumulation of reference sequences from identified specimens, not only in the light of "DNA barcoding", but also for data sharing between researchers (*e.g.* National Center for Biotechnology Information; https://www.ncbi.nlm.nih.gov/), leading to the creation of a variety of online databases, specifically designed for the taxonomic assignment of amplicon sequencing datasets, that are constantly updated by the registered users (*e.g.* BOLD, Ratnasingham and Hebert,

2007; UNITE, Nilsson et al., 2019), or by automated pipelines and by the guidance of specific taxonomic expert groups (*e.g.* Silva, Quast et al., 2012; PR2, Guillou et al., 2012). The user-updated databases, however, heavily rely on the correct assignment of a taxonomic identification performed by the user itself, and, for this reason, some of those projects require an extensive amount of metadata (*e.g.* primers used, taxonomic identifier, etc.) and additional information of the sequences uploaded (*e.g.* reference picture of the specimen), thus improving the validity of the reference sequences retrieved and detect possible misindentification or contamination cases.

However, many different taxonomic groups have been studied only recently using these techniques and the majority of the existing species, especially the least common species, do not have a corresponding sequence in the online database (Weigand et al., 2019), thus often leading only to a partial successful identification of the majority of sequences found in metabarcoding datasets. In order to bypass this issue, many bioinformatic tools have been developed to assign an identification to higher taxonomic levels regardless of the sequence representativity of the exact species represented in the metabarcoding datasets (Boyer et al., 2016; Munch et al., 2008), but still, the taxonomic library completeness represents one of the main obstacles to a thorough evaluation of the taxonomic composition in biological communities studied through “DNA metabarcoding” (Taberlet et al., 2018). Moreover, many different regions on Earth have been studied with a different sampling effort, leading to areas whose biological communities are scarcely represented on sequence reference databases (Weigand et al., 2019). The remoteness and isolation from other, more investigated regions, would certainly magnify this issue, and the Antarctic continent and the Southern Ocean represent evident examples.

A direct, simple search on the Web of Science (https://apps.webofknowledge.com, March 19th, 2021) conducted using the query “("Antarctic" OR "Antarctica" OR "Southern Ocean") AND ("amplicon sequencing" OR "metabarcoding" OR "DNA metabarcoding" OR "eDNA" OR "Environmental DNA" OR "HTS" OR "High Throughput Sequencing")” in the “Topic” field reported a total of 180

papers published from 2000 to 2020 with ~79% of the total publications occurring in the last 4 years (2016-2020). This shows how studies on Antarctic biodiversity conducted using such methodologies are still limited in number, as appears in the data shown in the recent review on “DNA metabarcoding” by (Compson et al., 2020). As an example, Brannock et al. (2018) reported that at 2018 only two papers were published on Antarctic meiobenthic communities using “DNA metabarcoding” analyses. Nonetheless, the potentialities of “DNA metabarcoding” have been already tested in both the Southern Ocean and continental Antarctica, indicating, for example, its usefulness in discriminating spatial turnover based on environmental characteristics and detecting the well known vertical migrations of copepods (*e.g.* Czechowski et al., 2016; Suter et al., 2020). These potentialities are particularly useful in monitoring biological communities, especially in light of the increasing threat of biological invasions due to climate change (Holland et al., 2021).

Marine protected Areas (MPAs) are an important management tool that can be used to protect, maintain, and restore biodiversity and ecosystem services. Different MPAs have been established in the Southern Ocean, in both the continental and sub-Antarctic regions (Brasier et al., 2018; Fabri-Ruiz et al., 2020). The MPA established in 2016 in the Ross Sea still represents the world's largest marine reserve to date. It includes, besides the area bounded by the General Protection Zone, a “Krill Research Zone” and a “Special Research zone” adjoining the former (CCAMLR, 2016). The region has been protected by both its remoteness and harsh weather conditions (Brooks et al., 2019), includes four ecoregions (of those evidenced by Douglass et al., 2014) and several environmental types, ranging from the Ross and Oates continental shelves to the Pacific abyssal plain, from the Scott Seamounts to the Balleny Islands (from data in Douglass et al. 2014), making it particularly relevant for the conservation of Antarctic communities.

However, since its establishment, no systematic research on the effects of this event has yet produced any defining result, and, more importantly, no reference baseline is available to perform it. Moreover, the current network of Antarctic MPAs (thus comprising both the Ross Sea and South Orkney Islands

southern shelf MPAs) do not provide a sufficient level of representativity of the total environmental types and ecoregions that can be identified in Antarctica, especially in light of the future predicted environmental changes (Fabri-Ruiz et al., 2020), with all the consequences that a reduced connectivity between distant but similar environmental types can have on the resilience of biological communities (Fabri-Ruiz et al., 2020). In this context, the additional information provided by genetic material can increase our understanding of the structure and connectivity between biological populations in different areas of the Southern Ocean and, thus, improve the development of an effective planning of MPA networks (Jenkins and Stevens, 2018). The development of bio-monitoring programs performed using HTS methodologies at a continental scale could also provide the right instruments to evaluate the effects of these MPA networks on the connectivity between populations in the long-term, thanks to the potentialities of metabarcoding datasets applied to the study of intra-specific patterns and phylogeographic features for hundreds of species simultaneously (*e.g.* Turon et al., 2020).

Terra Nova Bay (TNB) is a ~70 kilometers long inlet, lying between Cape Washington and the Drygalski Ice Tongue along the coast of Victoria Land, in eastern Antarctica (https://data.aad.gov.au/aadc/gaz/scar/). Since its discovery, which took place during the British National Antarctic Expedition (1901-1904), this area has been extensively studied only after the establishment of the Italian research base “Terra Nova Station”, later called “Mario Zucchelli Station” (MZS), in 1985 and the first Italian oceanographic expedition that was conducted from 1987 to 1988 (Faranda et al., 2000). The Italian research base MZS is located approximately at the centre of the bay and provides facilities and support for 85 people on average (between research and logistic personnel) operating only during the Austral summer from mid-October to the beginning of February.

The first Italian Antarctic expeditions were conducted in a poorly studied region, with only a limited amount of information and data available on all aspects of scientific research (Amato, 1990; Faranda et al., 2000). Notwithstanding the fact that important aspects on the structure and dynamics of both

the benthic and planktic communities in TNB have been disclosed (*e.g.* Cattaneo-Vietti et al., 2000), no long-term bio-monitoring program has been conducted to date, and, furthermore, scientific activities conducted on these communities using HTS methodologies are rare and almost exclusively focused on prokaryotic communities (Giudice and Azzaro, 2019).

The Italian National Antarctic Museum (MNA) was established in 1996 with the specific intent to preserve, study and make available to the scientific community all the material collected in Antarctica by the Italian National Antarctic Program (PNRA) scientific activities. This institution has an interuniversity organization, with different universities hosting specific typologies of samples, with the section hosted by the University of Genoa specialized in preserving biological samples (Schiaparelli et al., 2018). Since its establishment, the section of Genoa of the MNA acquired more than 10,000 biological samples and has continuously contributed to the major repositories of species occurrences such as the Global Biodiversity Information Facility (GBIF, 2021), through a variety of publications (Cecchetto et al., 2019, 2017; Giuseppe Garlaschè et al., 2019; Ghiglione et al., 2018; Piazza et al., 2014).

The main objective of this thesis is to provide the first quantification of DNA reference library completeness in Antarctica, by creating a bioinformatic pipeline that gathers all the available sequences from the major genetic data repositories, and focusing on the most common metazoan phyla occurring in the Ross Sea MPA, here used as a case study. Moreover, the structures and temporal dynamics of biological communities from shallow waters of TNB have been investigated using "DNA metabarcoding", which, to my knowledge, are the first studies of this kind conducted in this area using this kind of methodology, highlighting the advantages, issues and possible future implementations.

The first chapter of this thesis will cover the development of the bioinformatic pipeline used to create the first DNA barcode reference library for the metazoan species occurring in the Ross Sea MPA, by

gathering all the Cytochrome c oxidase subunit I (COI) sequences available in the major public repositories (*i.e.* NCBI and BOLD). The structure and functioning of the bioinformatic pipeline used here will be thoroughly described. Considerations of the knowledge gap between species occurrences in the area and DNA barcode representativity will follow, and the created Ross Sea reference library will be "tested" using a metabarcoding dataset (whose creation and refinement will be discussed in Chapter 2), in order to evaluate the quality and applicability of the assembled library.

The second chapter will discuss the application of "DNA metabarcoding" to the study of macrozoobenthic communities that colonized six artificial structures deployed from 2015 to 2018 in TNB. A preliminary evaluation of the community composition and development over a period of three years will be performed and confronted with the previous knowledge on growth of benthic pioneering communities in Antarctica. Finally, alpha diversity metrics will be computed and compared for both the data presented here and a dataset from another publication that employed the same analyses performed on the same kind of artificial structures deployed outside of the Southern Ocean, in temperate and tropical areas.

The third and last chapter of this thesis will describe the application of "DNA metabarcoding" on the filters of the MZS' desalination plant, in order to disclose the temporal dynamics of nanoplanktic and particle-attached bacterioplanktic communities in TNB. The applicability of HTS methodologies will be assessed using a series of filters collected in January 2012 and 2013. Intra-annual dynamics on the Antarctic coastal nanoplankton were disclosed, highlighting the importance of extreme, stochastic events such as katabatic wind pulses, which triggered dramatic, short-term shifts in coastal nanoplankton composition. This chapter was published in Cecchetto et al. (2021).

Abbreviations and conventions/specifications:

Some of the terms adopted in this thesis are reported in the form of abbreviations or acronyms that will be here listed with the corresponding full words. Nonetheless, the full words of most of these abbreviations will be reported in the first statement recurring across the manuscript. Some formatting conventions have been adopted to discriminate the names of the scripts, programs and commands that were created by the PhD candidate or used in the analysis. The names of the scripts that were created by the PhD candidate are reported in **bold**, whereas those of all the other programs, R packages and scripts used in the analyses, but created and published by other authors, are reported in *italic*. Whenever I refer to the names of specific commands and options, those are reported in double quotes "" and formatted in *italic* as well.

Images and Tables' numeration refer to the chapter they recur into, which means that multiple "figure 1", for example, are recurring in the entire thesis (*i.e.* one for each chapter). This was done in order to maintain a certain simplicity in the text, without recurring to a complicated caption numeration. Nonetheless, all figures and tables are reported in different folders at the online private Mendeley Data resource (https://data.mendeley.com/datasets/pj6gp5d758/draft?a=6d039418-0262-4b7e-b4da-c5bd9e9bf1ee), together with all the Appendix tables, figures and scripts.

The term "DNA metabarcoding" was introduced by Taberlet et al. (2012), originally referred to the "high-throughput multispecies (or higher-level taxon) identification using the total and typically degraded DNA extracted from an environmental sample" including also the "species identification from bulk samples of entire organisms, where the organisms are isolated prior to analysis", and derives from the DNA amplification and sequencing of targeted nuclear and organellar DNA regions of single specimens, which, if applied in a context of biodiversity and/or biological evolution research, and especially for species delimitation purposes, can be identified as "DNA barcoding", a term (and methodology) introduced by Hebert et al. (2003).

Recently, in order to reduce the inconsistencies that derive from misinterpretations of the terminology, an attempted standardization and clarification has been advocated for the term “environmental DNA” or “eDNA” (Pawlowski et al., 2020). This term would be used in the broadest sense, thus including both the “organismal DNA sourced from whole individuals most likely alive at the time of sampling” and the “extra-organismal DNA that can come from biological material shed from an organism” (Rodriguez-Ezpeleta et al., 2020). The right discriminations between the different applications of “DNA metabarcoding” in “environmental DNA” studies should simply be reported with a clear statement and description of the purposes, target organism and laboratory protocol applied to each particular study (Rodriguez-Ezpeleta et al., 2020). Thus, the right term to use for the analyses performed here would probably be “eDNA metabarcoding”, however, as all the analyses were conducted on bulk samples, targeting large, diversified but nonetheless well defined communities (nanoeukaryotic, particle-attached bacterioplankton and macrozoobenthic metazoans) and conducted to explore the taxonomic composition from *in situ* living communities actively sampled, I preferred to maintain the term “DNA metabarcoding” as originally intended by Taberlet et al., (2012).

Most sampling activities performed in the field and described in this thesis were conducted in the past by the Supervisor of the PhD candidate, and all the samples were stored at the section of Genoa of the Italian National Antarctic Musem (MNA). Nonetheless, all the analyses were performed by the PhD candidate during the three years of the PhD course (2017-2020), except for the DNA amplification and sequencing that were conducted by two different sequencing services providers, *i.e.* the Canadian Center for DNA Barcoding (CCDB, https://ccdb.ca/) and IGA Technology (https://igatechnology.com/).

Acronyms/Abbreviations	Definition
ARMS	Autonomous Reef Monitoring Structures
AWS	Autonomatic Weather Station
BOLD	Barcode of Life Data System
CAML	Census of Antarctic Marine Life
CCAMLR	Convention for the Conservation of Antarctic Marine Living Resources
CCDB	Canadian Centre for DNA Barcoding
COI	Cytochrome c oxidase subunit I
GBIF	Global Biodiversity Information Facility
HTS	High Throughput Sequencing
MNA	“Museo Nazionale dell’Antartide” - Italian National Antarctic Museum
MPA	Marine Protected Area
MZS	Mario Zucchelli Station
NCBI	National Center for Biotechnology Information
NMDS	non-Metric Multidimensional Scaling
OTU	Operational Taxonomic Unit
PNRA	“Programma Nazionale di Ricerche in Antartide” - Italian National Antarctic Research Program
POC	Particulate Organic Carbon
RSRL	Ross Sea Reference Library
SSTr	Sea Surface Temperature range
TNB	Terra Nova Bay
WoRMS	World Register of Marine Species
ZOTU	Zero-radius Operational Taxonomic Unit

Chapter 1:

Quantifying the gap between species occurrences and COI sequence coverage for Metazoa: an automated pipeline for the assembly of a DNA Barcode reference Library in the Ross Sea MPA.

1.1. Aims of the study:

1. Download all COI sequences available in the major genetic data repositories for the metazoan species reported by GBIF in the Ross Sea MPA;
2. Document the characteristics and overall quality of information gathered by multiple genetic data repositories (public as BOLD and NCBI and private as the MNA), and use the recovered data to quantify the gap between species occurrences and available barcode sequences for the area investigated;
3. Test the efficiency of the recovered data on an Antarctic metabarcoding dataset and compare the results with an up-to-date generic DNA barcode reference library (MIDORI, Machida et al. 2017)

1.2. Introduction

The study of Antarctic communities through the application of molecular techniques has received increasing attention in the last decade. Since the International Polar Year 2008/09, several international campaigns and expeditions with a variety of objectives, from disentangling the cryptic diversity of circumpolar species (*e.g.* Hemery et al., 2012), to inspecting phylogenetic histories (*e.g.* Bogantes et al., 2020), have been organized and carried on. Some of these campaigns were organized under the umbrella of the CAML (Census of Antarctic Marine Life) Barcoding Campaign (Schiaparelli et al., 2013), for a total of 18 international oceanographic Antarctic expeditions, resulting in circum-Antarctic sampling from the shallow shelf to the deep-sea (Grant et al., 2011). In just a few years since the establishment of the CAML, numerous barcode sequences were added to public repositories, reaching fifty times the number of previously known available molecular data for DNA barcoding (Grant et al., 2011; Grant and Linse, 2009), mainly focusing on the Cytochrome c oxidase subunit I (COI) marker. However, no review on this topic has been realized recently, and the last update report on molecular studies for Antarctic marine invertebrates has been published several years ago (Riesgo et al., 2015).

Since then, many different research projects were completed and the corresponding publications further increased the available data (*e.g.* Brasier et al., 2016; Christiansen et al., 2018). In those papers, the known discrepancy in the available number of barcodes for different taxonomic groups, such as Porifera and Annelida, as already highlighted by Grant et al. (2011), was further remarked. A similar discrepancy was also recognized for different geographic locations, such as the Antarctic Peninsula, the Dumont D'Urville, Weddell and Ross seas, hosting the majority of barcode sequences (Grant et al., 2011). The increasing availability of public and accessible molecular data not only favors the recognition and taxonomic assignment of unknown organisms, in the form of metazoan DNA barcoding, as originally proposed by Hebert et al. (2003), but also facilitate the understanding of evolutionary and ecological processes behind the structure of current biological communities,

especially in peculiar environmental settings such as those characterizing the Southern Ocean (Allcock and Strugnell, 2012).

In this context, one of the most useful instruments we have for gathering all the available data is the development and update of DNA barcode reference libraries. Many different public repositories are continuously updated and provide a collection of identified sequences for a variety of different taxa, markers and ecosystems (see Taberlet et al., 2018 for a general review). The need for high quality reference libraries increased with the advent of metabarcoding research (Taberlet et al., 2012), which heavily relies on their completeness and overall quality (Fontes et al., 2021). Many different DNA barcode reference libraries are realized by combining data from different data repositories in order to represent as much diversity as possible, aggregating a variety of sequences belonging to different taxonomic groups and without discrimination on the geographical area of origin (*e.g.* Machida et al., 2017). These kinds of libraries are especially useful when the information for the area investigated is scarce and/or the taxonomic assignment at the species level is not a priority of the study, thus representing an optimal application for preliminary analyses.

However, more complex metabarcoding studies require more sophisticated analyses, thus the employment of more refined DNA barcode reference libraries. In this sense, the restriction to specific taxonomic groups, with known distribution for the area investigated, would decrease the recovery of false positive taxonomic assignments, especially if this filtration is adopted at the species level (Questel et al., 2021). Thus, the production of DNA barcode reference libraries would greatly benefit from the integration of different kinds of information from multiple data repositories, especially considering the aforementioned need to develop high quality data, refined for specific geographic areas, taxa, etc. A multitude of tools are continuously developed for the creation of reference libraries, some of them adopting information from different biodiversity databases as a cleaning or standardizing procedure (*e.g.* the application of WoRMS data as a “contaminants” filter in Arranz et al., 2020).

In the last decades, species occurrence records from online databases such as the Global Biodiversity Information Facility (GBIF: The Global Biodiversity Information Facility, 2021; https://www.gbif.org/), became an indispensable resource for numerous research topics (GBIF Secretariat, 2021). These databases are constantly updated with new datasets, studying different taxonomic groups and geographic locations and following standardized procedures for the uploading of new occurrences (Penev et al., 2017), allowing the aggregation of high quantities of data in the same standardized format. However, the high number of occurrences gathered constantly by these repositories have different origins (defined, for example, by the Darwin Core Term "*basisOfRecord*", Wieczorek et al., 2012), meaning that they can refer, for example, to museum vouchers ("*preservedSpecimen*"), originate from old literature ("*HumanObservation*") or even correspond to media recordings such as camera traps photos ("*MachineObservations*"). The difference in record origins may inevitably vary the meta-data requirements each record must present to be accepted in the public repository, showing discrepancies in the reliability the resource downloaded (Zizka et al., 2019), as well as in their additional information. For this reason, gathering data from these public repositories requires a thorough cleaning process aimed at reducing as much as possible all the errors that may be occurring in the used resource.

These discrepancies hamper the ability to reproduce fast and reliable meta-analyses, which are inevitably reduced to specific taxonomic groups, geographic areas, datasets etc. (Gratton et al., 2017). These discrepancies are further magnified if we consider different data repositories (genetic, biological trait, etc.), often providing a different amount of available information for the same species. Recently, a study managed to estimate the overlaps between the most comprehensive sources of geographic, genetic and trait-based botanical data globally, showing that only less than 18% of the world's plant species share available and public knowledge on each different data repository (Cornwell et al., 2019). In fact, the reduced amount of additional, complementary data associated to public DNA sequences is a known issue ("A place for everything", 2008), and this is particularly true

when we refer to precise geographical information (Gratton et al., 2017), not only in public repositories, but also in publication's metadata (Pope et al., 2015). Different initiatives tried to address this issue by requiring particular data standards to be met upon the acceptance of sequence data, for both publications (Sibbett et al., 2020) and databases (*e.g.* the BOLD System, Ratnasingham and Hebert, 2007). Nonetheless, occurrence data repositories such as GBIF can still be considered the main source for species distribution, especially in remote and/or poorly sampled areas such as Antarctica and the Southern Ocean in general (Guillaumot et al., 2018). Providing a connection between all of these information sources is a crucial step towards the full automation of barcode reference library assembly.

1.3. Materials and Methods

Data assembly and cleaning was performed using a bash script (**main_NCBI_retrieval.sh**) interacting with two R scripts (**bold_retrieval.R** and **worms_check.R**), whereas the efficiency test on the recovered data was performed using a bash script (**obitools_arms.sh**) followed by an R script (**ARMS_barplot.R**). The first section of the Materials and Methods (section 1.3.1.) will briefly describe the taxonomic coverage and the geographic limits of the area investigated for the creation of the DNA barcode reference library, the second (section 1.3.2.) will describe the origin of the supplementary sequences obtained at the MNA and incorporated in the analyses and the third (section 1.3.3.) will describe the bioinformatic tools adopted and the analyses performed during the library assembly.

1.3.1. Taxonomic coverage and area investigated

The area investigated spans between 150°E and 150°W of longitude and covers approximately 2 million square kilometers of the Ross sea, from the Ross continental shelf to the abyssal plains on the western region and the Oates continental shelf and Balleny Islands on the eastern region, with the inclusion of the northern Scott seamounts and a portion of the Pacific Antarctic Ridge between 163 and 168°E and 60 and 62,30°S (Fig. 1).

The taxonomic groups investigated referred to 15 phyla, most of which are predominantly represented by benthic species (Appendix Tab. 1). These groups ranged from Phylum to specific classes, in order to avoid the recovery of records belonging to unwanted groups (*i.e.* not all Arthropoda have been recovered, but only some specific classes, thus excluding unwanted groups such as Insecta or Arachnida).

The class Scyphozoa was also included, notwithstanding the fact that its Antarctic species are predominantly holoplanktic and only a few species of the genus *Atolla* Haeckel, 1880 and *Periphylla* F. Müller, 1861 are characterized by benthic life stages (Brandt et al., 2007).

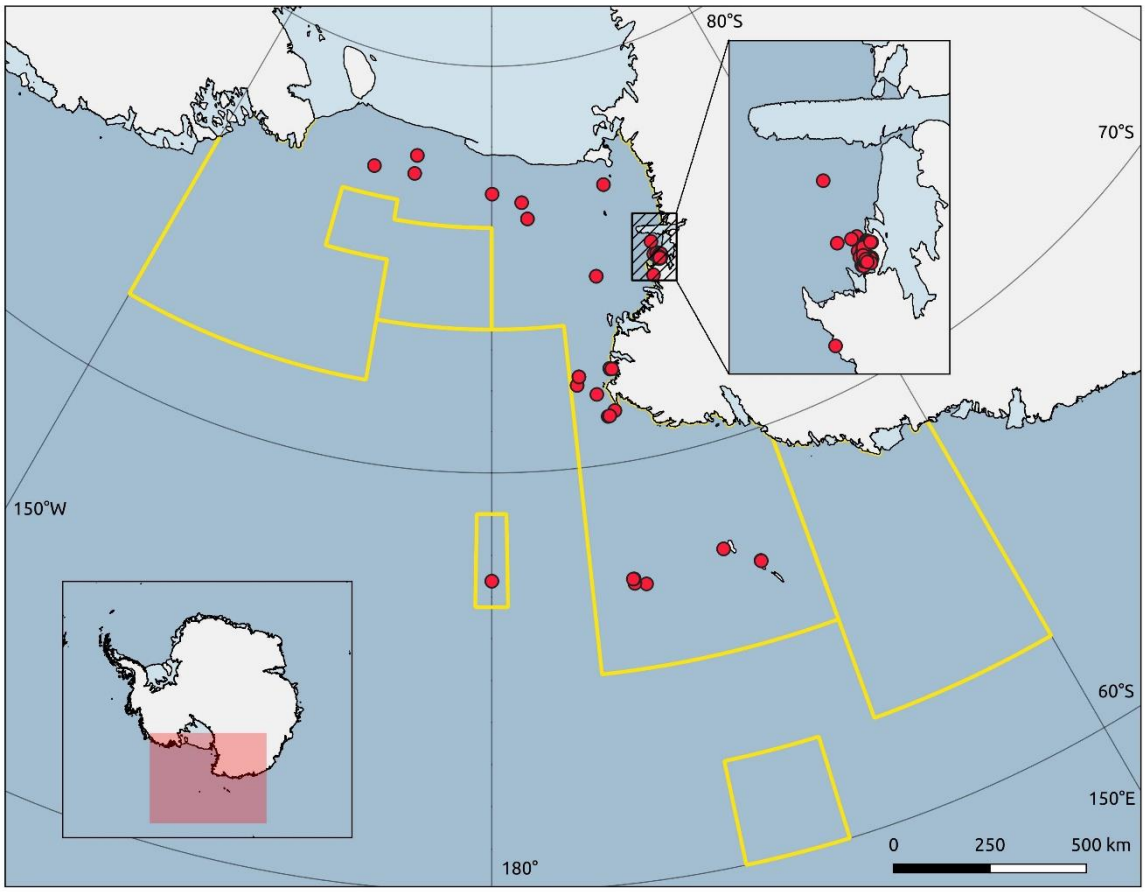


Figure 1: Extension of the different Protected Zones of the Ross Sea MPA (yellow). Red dots indicate the sampling location of the MNA samples included in this study, with an overview on Terra Nova Bay, on the upper right corner, where most of the samples were obtained.

1.3.2. Samples choice and production of DNA Barcodes from the MNA

Several hundreds of specimens were selected at the MNA for the production of COI sequences, for a total of 940 samples belonging to 7 phyla (Annelida, Arthropoda, Bryozoa, Chordata, Echinodermata, Mollusca and Porifera) and 15 classes (Actinopterygii, Bivalvia, Gastropoda, Polyplacophora, Demospongiae, Polychaeta, Gymnolaemata, Stenolaemata, Asteroidea, Holothuroidea, Echinoidea, Ophiuroidea, Crinoidea, Malacostraca, Pycnogonida; See Appendix Tab. 2 for details on each specimen).

These specimens were sampled during two New Zealand expeditions (National Institute of Water and Atmospheric Research, NIWA) and 8 Italian Expeditions (Italian National Antarctic Program, PNRA), from 2002 to 2019. All samples were preserved in ethanol or at -20°C and are currently

stored at the same conditions, except for a small quantity of specimens (approximately 1,5%) now stored in dry conditions or on a golden coated STUB.

All the samples were processed at the Canadian Center for DNA Barcoding (CCDB – University of Guelph, Canada). Briefly, a portion of tissue corresponding to approximately 1 cubic millimeter was sampled from each individual and placed in a 8×12 well microplate filled with absolute ethanol. The microplate was then shipped to the University of Guelph and processed following the CCDB automated standard protocols for extraction, amplification and sequencing (http://ccdb.ca/resources/). The primers adopted for the PCR amplification are listed in table 1.

Forward / Reverse Primer Codes	Citation
BivF4_t1 BivR1_t1	Layton, 2012
C_FishF1t1 C_FishR1t1	Ivanova et al., 2007
C_GasF1_t1 GasR1_t1	Stein et al., 2013
C_LepFolF C_LepFolR	Hernández et al., 2014
dgLCO-1490 dgHCO-2198	Meyer, 2003
EchinoF1 HCO2198	Ward, Holmes and O'Hara, 2008 Folmer et al., 1994
LCO1490 HCO2198	Folmer et al., 1994
LCO1490_t1 HCO2198_t1	Foottit et al., 2009
LCOech1aF1 HCO2198	Layton et al., 2016
polyLCO polyHCO	Carr et al., 2011

Table 1: Primers adopted for the amplification of the MNA samples and corresponding citation

1.3.3. Pipeline structure, bioinformatic tools and data analyses

All the analyses were conducted on the same laptop (16Gb of memory and an Intel® Core™ i7 processor) and performed using different scripts, two in bash and the others in R (Fig. 2). The first script (**main_NCBI_retrieval.sh**) acts as the main body, and the entire analyses can be divided into 4 major steps:

1.3.3.1. Species list preparation

The list of chosen taxonomic groups (Appendix Tab. 1) was queried on the GBIF database using the R package *rgbif* (Chamberlain et al., 2021; Chamberlain and Boettiger, 2017). All the occurrences belonging to the taxonomic id key of each group were downloaded only if reported as “preservedSpecimen” and with geographical coordinates occurring inside the area bounded by the Ross Sea MPA (https://gis.ccamlr.org/home/ccamlrgis) extracted from the Quantarctica dataset collection (Matsuoka et al., 2018) in QGIS (QGIS Development Team, 2020, version 3.4) in WKT (Well Known Text) format. Some records were reported on GBIF at a lower taxonomic level then the Species (*i.e.* with the string "SUBSPECIES" and "VARIETY"). For those records the script checked if other records retrieved were reported with the same species they belonged to, and only that taxonomic level was used for downstream analyses.

Next, an interactive session in R was opened and all the species names retrieved from GBIF were searched on WoRMS (Horton et al., 2021) using the R packages *taxize* (Scott Chamberlain and Eduard Szocs, 2013) and *worrms* (Chamberlain, 2020a). The interactive command “*get_wormsid*” (in *taxize*) was set in order to return a successful match only when the searched name was labeled as "accepted" and “Marine” or “Brackish” on WoRMS, thus discriminating synonyms or alternate representation and non-marine species. The names that did not return a successful result were searched again using the command “*wm_records_taxamatch”* (in *worrms*), whose “fuzzy search” method resulted more flexible than the one adopted by the “*get_wormsid*” command, especially for those scientific names which included the subgenus in its official form. If again no corresponding name at the Species level

was found, those names were reported as "not found" and not used for downstream analyses. For all the “accepted” names the taxonomic classification from Phylum to Species was retrieved by searching on WoRMS the corrisponding correct AphiaID using the *“classification”* command of *taxize*.

1.3.3.2. Data retrieval

In the event that a particular name was not updated to the version accepted by WoRMS, both the “unaccepted” versions and the “accepted” one were searched in the Taxonomy database of the NCBI (Agarwala et al., 2016), thus increasing the following records retrieval in the Nucleotide database. Each species was searched in the NCBI Taxonomy Database to recover all the corresponding TaxId (NCBI Taxonomy Identifier) codes, together with the taxonomic lineage corresponding to that species (namely, the Phylum, Class, Order, Family and Genus names). The TaxId was then searched on the Nucleotide Database (Agarwala et al., 2016) to retrieve all the records corresponding to it. If the record included one of the INSD (International Nucleotide Sequence Database Collaboration - http://www.insdc.org/documents/feature-table) features keys “CDS”, “gene” or “product”, both the accession-number and the corresponding feature key value were downloaded and then filtered to include only those corresponding to the COI marker (*e.g.* “COX1” for the “CDS” feature key or "cytochrome oxidase subunit 1" for the “product” feature key). The filtered accession-numbers were then queried again in the NCBI Nucleotide Database to retrieve both the specimen record information (latitude and longitude, PCR primers, collection date etc.) and the sequence in fasta format. All of these operations were performed using a combination of different programs of the 14.1 version of the Entrez Direct suite (Kans, 2020), namely the *esearch*, *efetch*, *efilter* and *xtract* programs.

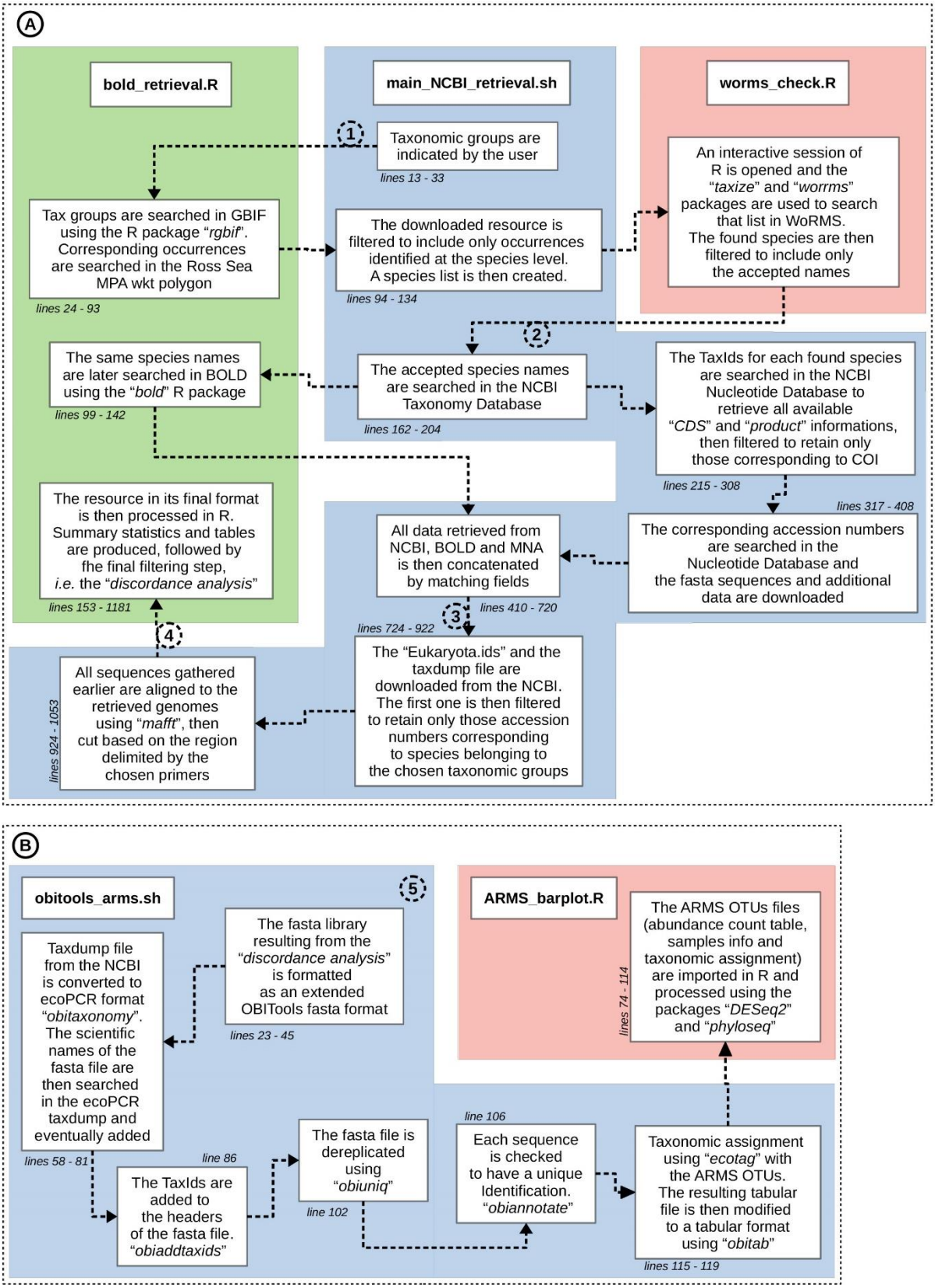


Figure 2: Pipeline structure of all the major analyses performed for this chapter. a) Analyses performed for the data retrieval from NCBI and BOLD; b) Analyses performed for the taxonomic assignment test on the ARMS OTUs. Numbers in dashed circles correspond to subchapters of the third section of Materials and Methods.

The same list of species names was then queried again on the BOLD System Database using the R package *bold* (Chamberlain, 2020b) and more specifically the "*bold_seqspec*" command, which returns both the specimen record information and the nucleotide sequence in tsv (Tab Separated Values) format. The returned data-frame was then saved and filtered to remove all those records labeled as "Mined from GenBank, NCBI" and retain only those corresponding to the "COI-5P" marker. The ProcessId from the retrieved BOLD records were then searched in the retrieved NCBI records and eventually removed from the latter group in order to avoid duplication, in case some records were uploaded independently on both databases.

The private records obtained by the sequencing of the MNA samples were then merged with those retrieved from BOLD and the NCBI in a single tsv file, and the different fasta files concatenated.

1.3.3.3. Alignment with NCBI mitochondrial Genomes and extraction of the COI region

Next, the information on eukaryotic species with available assembled mitochondrial genomes was obtained from the ftp network of the NCBI Genome Database (https://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/IDS/Eukaryota.ids) and filtered to retrieve all the accession-numbers relative to those species belonging to the queried taxonomic groups (Appendix Tab. 1), using the taxdump file of the NCBI (from https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/new_taxdump.zip). Similarly to the NCBI search, all the "CDS" information for those accession-numbers were inspected and only those corresponding to the COI marker were retained. If the number of available genomes for a particular group was greater than 10, then only 10 genomes were chosen randomly and used for all the following analyses (see Appendix Tab. 3 for the accession-numbers of the genomes used, together with the COI region coordinates), including the performance testing on the metabarcoding dataset (see the next paragraphs for specifications). The specific coordinates for that marker were retrieved and each fasta sequence was trimmed using the "*faidx*" command of *samtools* (Li et al., 2009), thus retaining only the COI region. For each taxonomic group both the genomes and the retrieved sequences belonging

to the same group were aligned using the command line version of *mafft* (Katoh and Standley, 2013). Briefly, all the mitochondrial genomes retrieved for each group were aligned together using the "*--auto*" option, which automatically selects an appropriate alignment strategy. Next, the retrieved sequences were aligned to the genomes alignment using the "*--auto*" and "*--addfull*" options, the latter used to keep the original length of the added sequences, thus not modifying the sequence to fit the alignment (https://mafft.cbrc.jp/alignment/software/addsequences.html).

A fasta file of the two primers amplifying the Folmer region (Folmer et al., 1994) of the COI was then aligned to the aforementioned alignment and the positions of the primer's nucleotides were exported ("*--mapout*" option) and inspected: If the length of the region included between the last base of the forward primer and the first of the reverse primer approximately corresponded to the typical length of the Folmer region of the COI (~658 bp) then the alignment was cut on those position, thus retaining only the Folmer region. All the alignments produced in the execution of the script reported the expected length for the region using those primers.

The new length of each of the retrieved sequences was then recalculated by the last part of the **main_NCBI_retrieval.sh** script, and the records corresponding to those sequences which didn't align with the Folmer region, and thus returned a new length of zero base pairs, were discarded.

For the performance testing of the reference library using an Antarctic metabarcoding dataset, the same procedure was applied, with the exception of the primers used, in this case the Leray degenerate primers (Leray et al., 2013), and all the alignments returned the right length of the expected region (in this case ~313 bp).

1.3.3.4. Summary statistics, figures and tables production and "discordance analysis"

The species names of the retrieved records were compared to the accepted species list obtained from WoRMS (and corresponding to the GBIF occurrences) and the names and taxonomic lineages were uniformed between synonyms and "unaccepted" names, without removing the original taxonomic

information of each specific database (BOLD, NCBI and MNA) which may be useful in case of unresolved taxonomic issues. Some species names of the MNA samples were not found in the WoRMS species list, meaning that they may potentially correspond to new records for the area. The species names for those records were again searched on the WoRMS database using the *worrms* R package, similarly to the initial part of the script and uniformed to the accepted taxonomy (the small amount of samples for which no accepted name was found corresponded to species yet to be described, and added from the MNA samples, thus still unavailable on the WoRMS database).

The *tidyverse* collection of R packages was used (Wickham et al., 2019a), together with *cowplot* (Wilke, 2020) to create the sequence length histogram. The package *ggVennDiagram* (Gao, 2021) was used to create Venn diagrams for the record's origin (MNA, BOLD or NCBI) of each queried Phylum. Maps were created using the R packages *simple-features* (Pebesma, 2018) and *rnaturalearth* (South, 2017). Many scripts used in the production of these plots were adapted from Gwiazdowski et al. (2015).

The final fasta file was imported in R using the package *ape* (Paradis and Schliep, 2019) and filtered to retain only those sequences (and corresponding record information) with a "new length" greater than 500 bp. The DNAbin sequences were then converted as a DNAstringset object and aligned using the *decipher* package (Wright, 2016) and the default specifications of the "*AlignSeqs*" command. The distance matrix was calculated using the default specifications of the "*DistanceMatrix*" command and then all the sequences were assigned to 99% similarity clusters using the "*complete*" method of "*IdClusters*" command. If the records assigned to a cluster presented an identical taxonomic classification or were the only record assigned to their cluster, they were labeled as "concordant" or "singleton" respectively. If a discordance at any taxonomic level inside a cluster was detected, those records were flagged with the corresponding discordant taxonomic level ("phylum", "class", "order" etc.). This "discordance analysis" was employed following a similar approach adopted by Gwiazdowski et al. (2015), and the discordance plot was produced using the *rgl* package (Adler et

al., 2021). Differently from Gwiazdowski et al. (2015), no species delimitation method was applied, and the analysis was conducted with a clustering at a 99% similarity threshold only, in order to show the generic sequence diversity and provide an indication of possible misidentifications, contaminations and uncertainty in nomenclature, following Machida et al. (2017). Next, the GBIF species list retrieved and the final records table were confronted and the proportion of species with a high quality barcode was calculated.

1.3.3.5. Performance test on a metabarcoding dataset

The last part of the "discordance analysis" in the **bold_retrieval.R** script removed all the records flagged as discordant at a taxonomic level higher than the genus (*i.e.* from Family to Phylum) from the fasta file of the region extracted earlier using the Leray degenerate primers (Leray et al., 2013), thus obtaining a "fasta library" composed only by records with a good alignment of the Folmer region (and thus the Leray region, which correspond to the 3' end of the Folmer region) and not discordant at higher taxonomic levels, as suggested by Machida et al. (2017) as a cleaning procedure for DNA barcode reference libraries.

A bash script (**obitools_arms.sh**) was used to format the fasta library as an extended OBITools fasta (https://pythonhosted.org/OBITools/attributes.html) adding the species name and the entire, "worms accepted" taxonomic lineage to the headers of the fasta sequences. The OBITools suite of python programs (Boyer et al., 2016) was then used for the final cleaning steps and the taxonomic assignment.

The taxdump file of the NCBI was downloaded, converted to ecoPCR format (Ficetola et al., 2010) using the *obitaxonomy* program, all the species represented by sequences in the fasta library were queried against it and eventually added to the ecoPCR files using the same command. The corresponding NCBI species TaxId were added to the headers of the fasta library using the *obiaddtaxids* program. The fasta library was then dereplicated using *obiuniq* and all the unique

sequences were checked to ensure that each would have a unique taxonomic identification using the *obiannotate* program.

A metabarcoding dataset was used to test the performance of the fasta library in the taxonomic assignment. This dataset was obtained by the metabarcoding analysis on scrapings of the plates of 6 Autonomous Reef Monitoring Structures (ARMS, Leray and Knowlton, 2015) deployed at 20 meters of depth in the area before the Mario Zucchelli Station in Terra Nova Bay (Ross Sea) and all retrieved and analyzed during the realization of this thesis to study the colonization of Antarctic benthic communities on artificial structures. All the specification on the methodologies adopted and results obtained in the analyses of the ARMS are reported in the second chapter of this thesis.

A similar procedure was applied to the latest version of the MIDORI reference library (Machida et al. 2017, release GenBank 241 of December 2020 Leray et al. in prep.). This library was downloaded in *mothur* format (Schloss et al., 2009) and, after formatting the headers as an extended OBITools fasta (again by including the species name and the taxonomic lineage), each sequence belonging to each investigated taxonomic group was aligned to the same genomes used earlier with *mafft* and following the same procedures (*i.e.* using the Leray primers). The newly filtered fasta library was then processed with the same OBITools programs used earlier, with the exception of two additional steps performed with *obigrep* to retain only those sequences identified at the Species level and with a good taxonomic description at least at the Family level, as suggested by the OBITools tutorial pipeline (https://pythonhosted.org/OBITools/wolves.html).

The performance on the created Ross Sea reference library (hereafter RSRL) was evaluated by visually analyzing the results of a taxonomic assignment on the ARMS dataset OTUs (Operational Taxonomic Units) using the default specifications of the *ecotag* program, with both reference libraries (MIDORI and RSRL) used alternatively in the taxonomic assignment.

The output files in extended OBITools fasta format were then converted to a tabular format using the *obitab* program, and then manually edited by retaining only the "scientific_name" attribute, which indicates the lowest most reliable taxonomic name assigned by the program *ecotag*, for each OTU. The resulting file was then analyzed by the "Match taxa" tool in WoRMS, in order to retain the entire taxonomic lineage of each scientific name, then manually edited to fill the empty cells of those taxonomic names for which the *ecotag* program couldn't assign a valid scientific name with the assigned scientific name followed by "_unidentified". Taxa barplots were generated in R (script **ARMS_barplot.R**) using the *phyloseq* package (McMurdie and Holmes, 2013) from the original, not transformed, count table after collapsing together all the replicates in the respective samples using the *DESeq2* R package (Love et al., 2014). Different taxa barplots were realized for the two taxonomic assignment methods (MIDORI and RSRL) at both the Phylum and Species taxonomic levels.

Additional informations on the structure and functioning of the scripts used can be found in the comments included in the scripts (Appendix Scripts).

1.4. Results and Discussion

1.4.1. GBIF records summary and WoRMS taxonomy normalization

The *rgbif* download retrieved 41,115 occurrences (GBIF Occurrence Download, https://doi.org/10.15468/dl.aq96re), reported as "preservedSpecimen" and referring to 51 storing Institutions (Tab. 2). The number of species reported by GBIF accounted for a total of 2,002 different names, out of which only 1966 were "accepted" names in WoRMS and labeled as marine or brackish, whereas the total number of species queried on the NCBI and BOLD repositories, comprising both the accepted and unaccepted versions of the names, summed up to 2,004 (Tab. 2).

Phylum	N° GBIF occurrences	N° GBIF Species	N° Storing Institutions	N° WoRMS Accepted Species	N° Names queried
Annelida	3509	184	17	179	181
Arthropoda	11210	453	20	445	450
Brachiopoda	164	12	4	12	13
Bryozoa	2372	232	10	229	234
Chordata	2682	178	28	178	179
Cnidaria	3544	191	21	189	193
Ctenophora	12	2	2	2	2
Echinodermata	8488	256	21	256	262
Hemichordata	93	6	3	6	6
Mollusca	5036	285	22	281	287
Nematoda	1487	47	10	38	41
Nemertea	176	6	4	6	6
Porifera	2130	141	9	140	145
Rotifera	5	4	1	0	0
Sipuncula	207	5	6	5	5
Total	41115	2002	51	1966	2004

Table 2: Statistics on the records retrieved from GBIF and inspected by WoRMS.

The discrepancy between the number of names reported by GBIF and the number of names accepted by WoRMS is evident. The GBIF Backbone Taxonomy (GBIF Secretariat, 2019) allows GBIF to integrate name-based information from different resources, with the Catalogue of Life (Roskov et al., 2020) acting as a starting point for a regular update (see https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c#citation). The adoption of multiple sources by the GBIF Backbone Taxonomy suggests adopting specific and unambiguous taxonomic cleaning steps (Zizka et al., 2020)

in the production of a barcode reference library. Other filtering steps would reduce the presence of erroneous occurrences, and WoRMS would further help this cleaning process by labeling each taxonomic name with ecological attributes (*e.g.* Marine, Terrestrial). The recovery of all the metazoan records occurring in the Ross Sea MPA reported also different unusual records, certainly the result of erroneous coordinates, such as the pulmonate gastropod *Achatinella mustelina* Mighels, 1845 in the Balleny Islands (occurrenceID: http://n2t.net/ark:/65665/340f6e6cb-5243-4525-9bac-35567143d10e). In fact, approximately 1.5% of the total species names reported by GBIF for the area were labeled by the adopted WoRMS taxonomy standardization procedure as "not found", "not marine" or with a different taxonomic status than "accepted", "not accepted" or "alternate representation". Regarding the Rotifera, only four species are reported by GBIF for the area investigated, however they were all labeled by WoRMS as not marine, consistently with the literature review on Antarctic Rotifera indicating that all the Rotifera records for Antarctica come from temporarily deglaciated areas along the coast (Garlaschè et al., 2019).

As the manual editing of biogeographic data repositories is unpractical and scarcely reproducible, the adoption of a systematic cleaning and standardization of these resources is necessary.

1.4.2. Ross Sea barcode reference library statistics

A total of 11,076 records (Tab. 3) gathered by the different scripts passed the final "discordance analysis" filtering step which removed all the records with a discordance at high taxonomic levels (as a final cleaning procedure suggested by Machida et al., 2017). These records represented 252 families and 635 species, with the most diversified groups at the Family level being the Gastropoda (32 families), the Malacostraca (28 families) and the Demospongiae (20 families). A slight discrepancy in this pattern is revealed by looking at the groups with the highest number of species, being the Actinopterygii (81 species), the Malacostraca (70 species), the Demospongiae and the Ophiuroidea (both 51 species) followed by the Gastropoda (48 species).

Phylum	Class	N° records Total	N° records NCBI/BOLD	% MNA records	N° Families	N° Species	N° concordance/singleton clusters	N° discordant cluster Genus	N° discordant cluster Species
Annelida	Polychaeta	841	638	24	19	47	170	2	0
	Clitellata	5	5	0	1	3	4	0	0
Arthropoda	Malacostraca	1701	1635	4	28	70	215	0	3
	Pycnogonida	878	839	4	9	43	131	1	8
	Hexanauplia	625	625	0	15	30	83	0	2
	Ostracoda	80	80	0	2	3	22	0	0
Brachiopoda	Rhynchonellata	3	3	0	1	3	3	0	0
Bryozoa	Gymnolaemata	18	10	44	5	10	10	0	1
	Stenolaemata	1	0	100	1	1	1	0	0
Chordata	Actinopterygii	1266	1216	4	19	81	129	3	8
	Ascidiacea	309	309	0	10	23	33	0	1
Cnidaria	Anthozoa	152	152	0	10	32	16	4	2
	Scyphozoa	50	50	0	2	2	13	0	0
	Hydrozoa	34	34	0	4	5	16	0	0
Echinodermata	Crinoidea	1462	1437	2	4	7	23	1	0
	Ophiuroidea	692	553	20	11	51	141	1	8
	Echinoidea	342	285	17	7	16	20	0	0
	Asteroidea	296	294	1	10	34	74	2	4
	Holothuroidea	152	8	95	5	15	17	0	1
Hemichordata	Graptolithoidea	1	1	0	1	1	1	0	0
Mollusca	Cephalopoda	780	780	0	11	22	48	0	6
	Gastropoda	693	663	4	32	48	195	0	0
	Bivalvia	182	146	20	13	15	35	0	0
	Polyplacophora	20	18	10	3	5	13	0	0
	Scaphopoda	15	15	0	1	1	1	0	0
Nematoda	Chromadorea	9	9	0	2	3	4	1	0
Nemertea	Hoplonemertea	251	251	0	1	1	2	0	0
	Pilidiophora	54	54	0	1	1	6	0	0
Porifera	Demospongiae	140	87	38	20	51	45	4	4
	Hexactinellida	13	13	0	2	8	7	0	2
	Homoscleromorpha	1	1	0	1	1	1	0	0
Sipuncula	Sipunculidea	10	10	0	1	2	5	0	0
	Total	11076	10221	8	252	635	1484	19	50

Table 3: Statistics on the Ross Sea Reference Library (RSRL) and on the "discordance analysis" performed on it.

The “discordance analysis” reported a total of 1,484 clusters at 99% similarity, more than a double of the number of species retained (Tab. 3). The taxonomic groups with the highest number of 99% clusters were the Malacostraca (215 clusters), the Gastropoda (195 clusters) and the Polychaeta (170 clusters), corresponding to a substantially lower number of species (70, 48 and 47 respectively) and possibly reflecting a higher cryptic diversity, which is known at least for some specific groups (*e.g.* Brasier et al., 2016). The opposite situation is observed in the records belonging to the Demospongiae (45 clusters and 51 species) and the Anthozoa (16 clusters and 32 species) for which a higher number of species corresponded to a lower amount of 99% similarity clusters, possibly as an effect of the slow mitochondrial DNA sequence evolution for those groups (Huang et al., 2008; Shearer et al., 2002). Moreover, the Demospongiae were also one of the taxonomic groups with the highest number and percentage of records belonging to clusters discordant at higher taxonomic levels (Tab. 4), notwithstanding the relatively low amount of total records in the dataset. The discrimination between misidentifications and contaminations is difficult to achieve without a direct comparison of the specimens, however, considering that all these clusters grouped records belonging to the same class (*i.e.* Demospongiae, see “% discordant records within same Class” in Tab. 4) and the known complexity of sponge identification (van Soest et al., 2012), it may also reflect a higher risk of misidentification and uncertainty in the nomenclature, as evidenced by Vargas et al. (2015).

Class	N° pre-discordance records	N° high rank discordant records	% high rank discordant records	% discordant records within same Class
Demospongiae	167	27	16.17	100
Actinopterygii	1278	12	0.94	100
Gastropoda	695	2	0.29	100
Ophiuroidea	715	23	3.22	80
Echinoidea	349	7	2.01	50
Polyplacophora	24	4	16.67	0
Crinoidea	1555	93	5.98	0
Asteroidea	312	16	5.13	0
Holothuroidea	154	2	1.3	0
Bivalvia	183	1	0.55	0
Total	5432	187	3	

*Table 4: N**umber of high rank discordant records** refers to the records that were removed by the discordance analysis. **% discordant records within same Class** corresponds to the percentage of records with a discordance in the same Class*

Incongruence at a lower taxonomic level (*i.e.* genera and species) would require a more thorough cleaning procedure, as the chance of dealing with cryptic species and thus the risk of removing valid data is nonetheless high. A higher number of discordance at the species level was reported for some of those groups showing a high number of species (*e.g.* Actinopterygii and Ophiuroidea, with 8 discordant clusters, see Tab. 3 and Fig. 3), however, as said earlier, it is difficult to discriminate between the different possible origins for this kind of discordance. Nonetheless, a high number of clusters discordant at the genus level was found for those groups characterized by a number of species higher than the clusters (*e.g.* Anthozoa and Demospongiae, see Fig. 3), again possibly reflecting the interpretations suggested earlier for those groups.

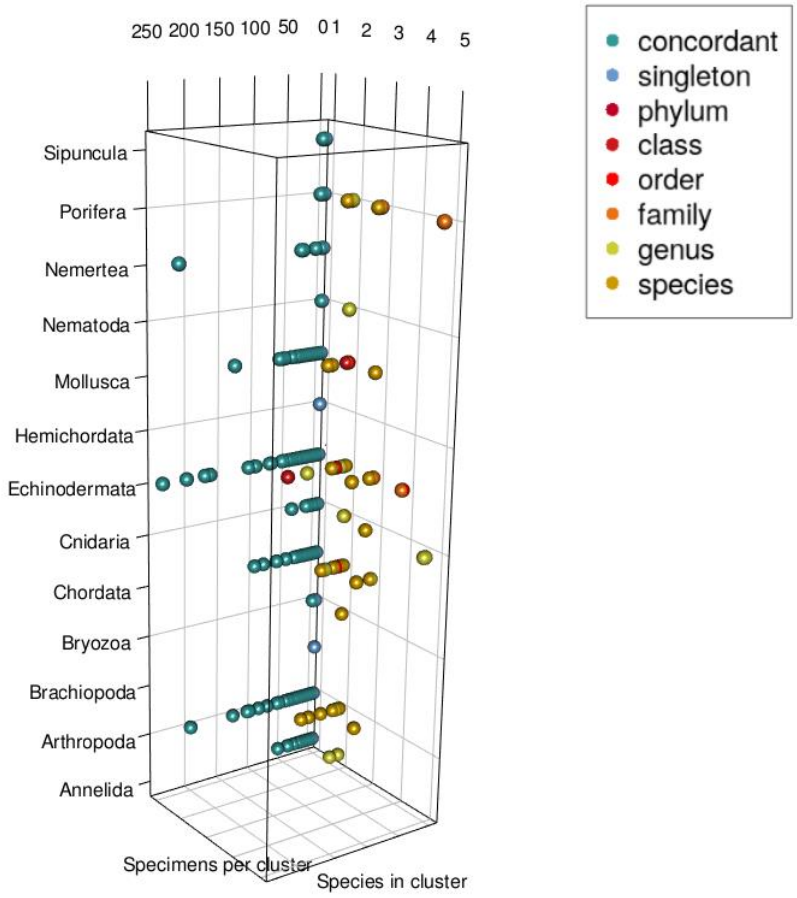


Figure 3: Discordance plot for all the retrieved phyla. Colors refer to the discordance taxonomic level. The plot shows both the number of species and the number of records in each cluster.

Most of the records were retrieved by the NCBI and BOLD repositories, with only a small portion (~8%) belonging to the additional sequences obtained from MNA samples (Tab. 3). Thus, for most of the taxonomic groups investigated, records were prominently or even exclusively gathered from these external resources, with the exception of some specific taxa, such as the Holothuroidea, the Bryozoa and the Demospongiae, covered by MNA records for the 95, 50 and 38% approximately and respectively (Tab. 3). These exceptions reflect to some extent the exclusion of certain ecologically relevant taxonomic groups from research campaigns dedicated to DNA barcoding and evidence the importance that even small, regional-specific collections still have in filling the gap between species occurrences and sequence coverage. This condition is even more exacerbated if we consider the number of new barcoded species that are added to the scientific knowledge, with ~71% of the Bryozoa and ~28% of Porifera species represented by sequences deriving from MNA samples (Fig. 4). Moreover, the high contribution of new species can be detected also for groups highly represented in the external resources such as the Arthropoda, Mollusca and Annelida with 3,182, 1,622 and 643 records from external sources respectively (see Tab. 3), for which respectively 12, 13 and 19% of the species are represented exclusively by sequences of the MNA collection (Fig. 4).

Another aspect to consider is the variability of sequence quality and characteristics of the different external resources. NCBI records hosts sequences from a variety of different projects and analyses (*e.g.* genome sequencing, other than amplicon sequencing), resulting in a higher variability in both sequence length, quality and genomic region analyzed, whereas BOLD records hosts mainly sequences of the Folmer region of the COI, with a lower variability in sequence length (Fig. 5a). As approximately 19% of the sequences retrieved showed a length of over 658 bp and comes mostly from the NCBI, this resource was identified as the one with the most diversification in that sense, a pattern that has generally been already recognized (Taberlet et al., 2018). Thus, the COI region trimming adopted after the genome alignment would evidently affect mostly the NCBI sequences,

with a substantial reduction in sequence length (Fig. 5b). On the other hand, the additional information provided by resources focusing on specific markers and taxa such as the BOLD (and others resources such as the UNITE database Nilsson et al., 2019) increase the overall sequence quality and validation, differently from the NCBI were this information is not required for validation and often neglected (Pentinsaari et al., 2020; Taberlet et al., 2018).

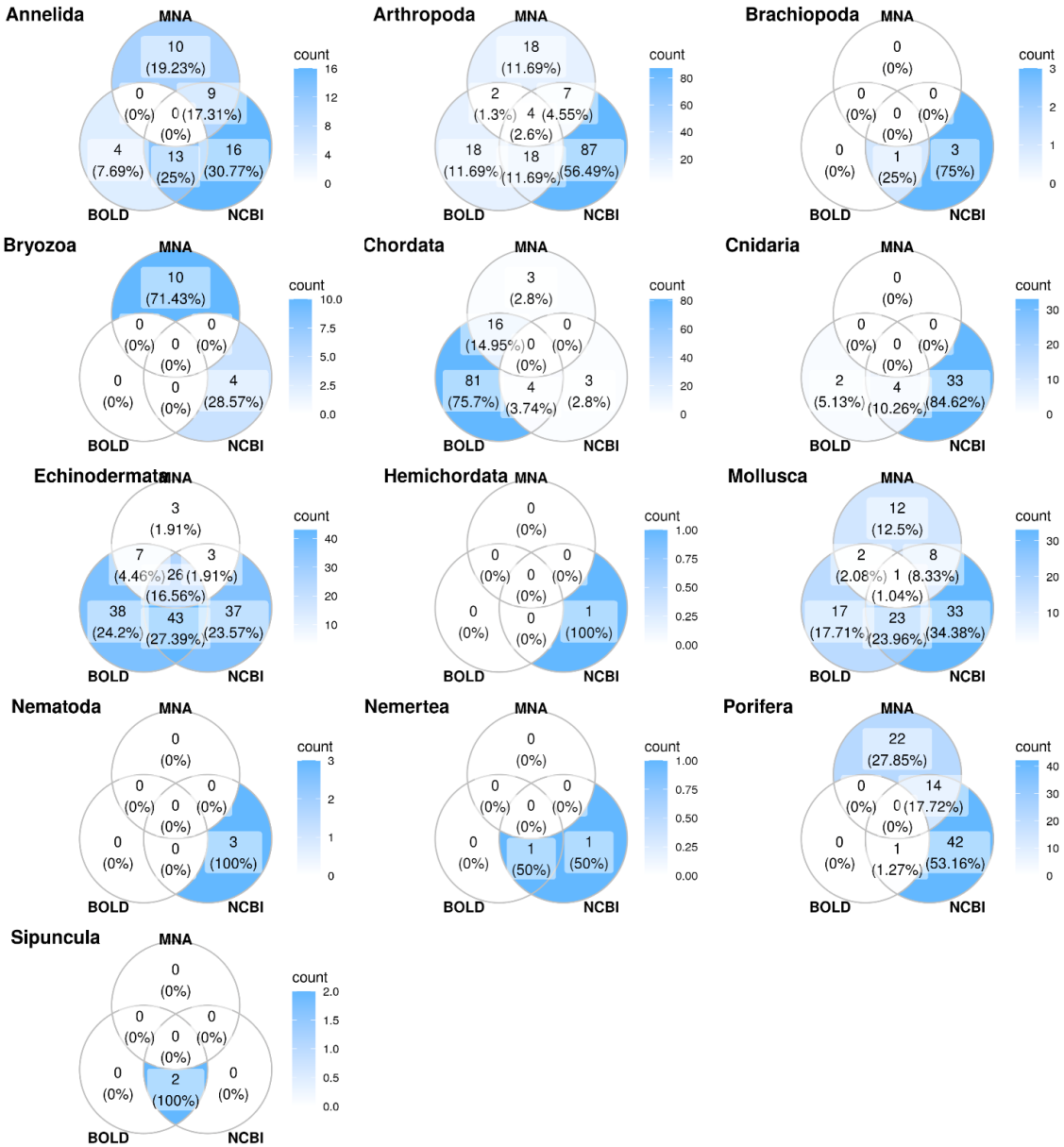


Figure 4: Venn Diagrams for all the phyla retrieved by the different repositories showing the number and percentage (in brackets) of all the unique and shared species.

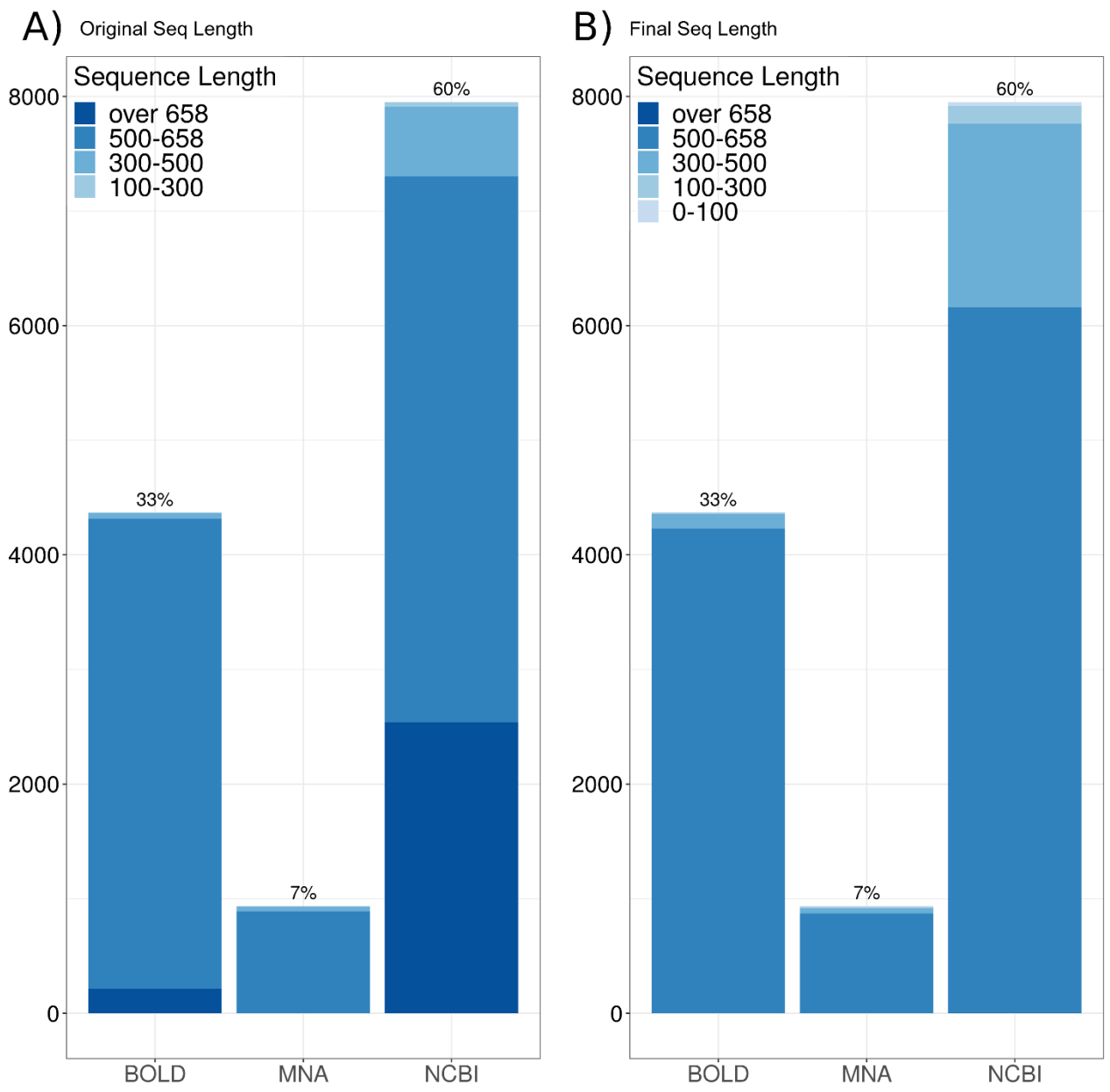


Figure 5: Barplot of the length distribution for the sequences retrieved by the different data repositories before (a) and after (b) the alignment with mafft and the extraction of the COI region.

For this reasons, the inclusion of local data, especially when provided with additional information can represent a substantial increase in the overall quality of the reference library. Here, as the MNA samples were processed at the CCDB and thus uploaded on BOLD, they contain the same quality of collateral data.

Regarding this topic, 7,162 records from the total 11,076 reported geographical coordinates in the collateral data (Fig. 6). The majority of them (~70%) were located inside the geographical limits of the Southern Ocean (< 60° South) and approximately 67% of the remaining ones were located in the area between the equator and the northern limit of the Southern Ocean (Tab. 5). Most of the Phyla with records having geographical coordinates were predominantly collected in the Southern Ocean geographic limits, with the exception of the Arthropoda, whose records were mostly collected between the equator and the northern limit of the Southern Ocean. By a close inspection of those records, the majority of them (550 out of 808) were identified as *Euphausia superba* Dana, 1850 and were uploaded on the NCBI with the same "authors" and "title" feature keys, thus possibly corresponding to the same oceanographic campaign (see accession number LC021725.1 for examples).

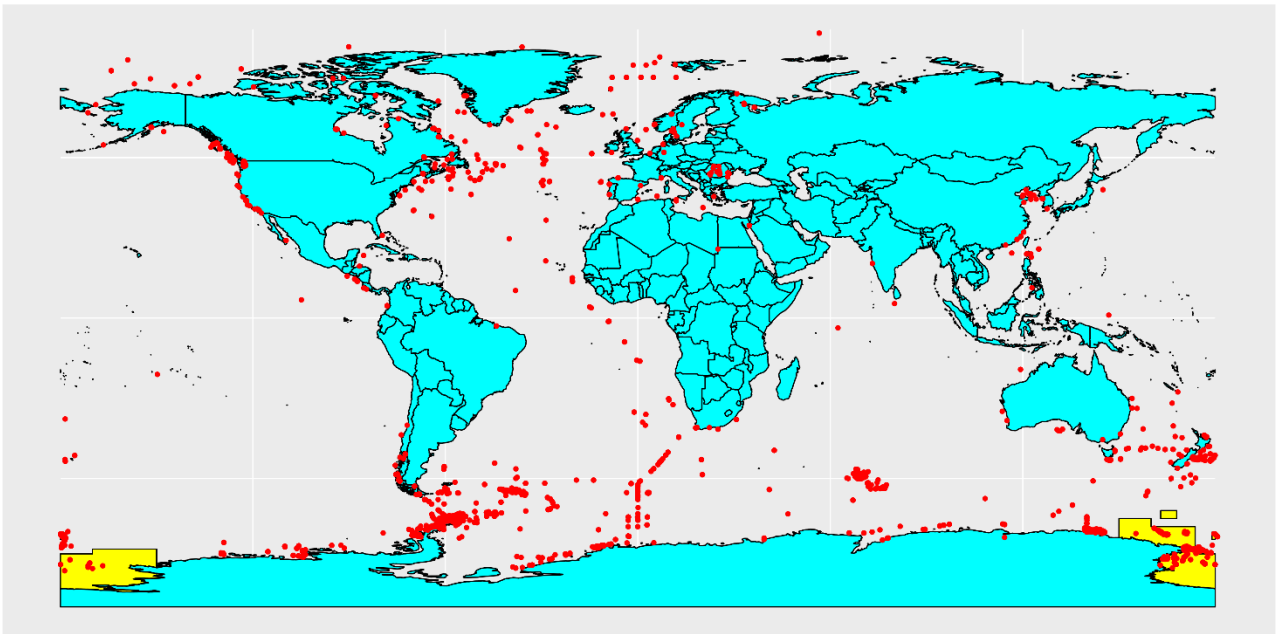


Figure 6: Geographic distribution of the retrieved georeferenced records (red dots). The polygon of the area investigated and queried on GBIF is shown in yellow.

Phylum	<60°S	>60°S to equator	from equator to 60°N	>60°N	% out South Ocean
Nematoda	0	0	1	0	100
Sipuncula	0	0	8	0	100
Cnidaria	2	3	44	4	96
Arthropoda	458	808	211	14	69
Nemertea	16	8	0	0	33
Mollusca	632	181	62	32	30
Annelida	517	43	117	27	27
Chordata	1131	154	129	9	21
Echinodermata	2214	229	35	0	11
Porifera	64	0	0	0	0
Bryozoa	9	0	0	0	0
Total	5043	1426	607	86	30

Table 5: Statistics on the georeferenced records.

By comparing the GBIF species list with the final library, the overall sequence coverage of the Ross Sea reference library corresponded approximately to 33% of the total species (Appendix Tab. 4). The highest values were recorded for specific phyla such as Chordata (~59%) and Porifera (~50%), followed by Echinodermata (~49%) and Sipuncula (~40%), the latter, however, only represented by 5 species. This condition is highly variable across different orders and classes of the same phylum, with the example of the Actinopterygii, reaching 75% of sequence coverage if we exclude the Ascidiacea (reaching only the 33% of sequences coverage). In the last review on molecular studies in the Southern Ocean (focusing exclusively on genetic diversity and connectivity) by Riesgo et al. (2015), Porifera and Annelida were identified as the least “barcoded” groups in Antarctica, with only a handful of species represented by sequences available in public repositories. The geographical specificity of this study won’t allow definitive considerations on the sequence coverage for these groups and for the entire Southern Ocean, however, these results suggest a turnaround of this condition, at least for Ross Sea. The overall sequence coverage observed here for the Ross Sea is inevitably also the result of different independent projects which provided an increasing number of sequences for different groups in the last decade (*e.g.* Brasier et al., 2016 for Polychaeta; Dettai et al., 2011 and Christiansen et al., 2018 for Actinopterygii and Vargas et al., 2015 for Porifera, including

data presented in this study), highlighting again the importance and remarkable results of such activities.

Nonetheless, as bio-geographical occurrences such as those from GBIF can't be treated as definitive and exhaustive checklists as such (*i.e.* without a proper and thorough inspection on the validity of each gathered dataset) these result would require further and more taxon specific investigations.

1.4.3. Taxonomic assignment performance on ARMS OTUs

The final fasta file, after the removal of the high rank discordant records, realignment with the Leray primers (Leray et al., 2013) and the OBITools quality filtering, counted only 4,194 sequences (Appendix Tab. 5). All the removed sequences were filtered out by the dereplication step in the OBITools pipeline, as the other quality improving steps were actually already performed in the previous pipeline (*i.e.* the *obiannotate* didn't filter out any additional sequence as all records already showed a unique identification). The taxonomic assignment performed using the *ecotag* program reported 18 OTUs, out of the total 224 from the ARMS sequence clustering (see Chapter 2 of this thesis for specifications), matching with an identity greater than 97% (thus possibly a valid match at the Species level) and without uncertainty in the taxonomic assignment due to multiple matches with sequences characterized by different taxonomies. Of the remaining OTUs, 17 were matched at the Species level, still without uncertainty in the taxonomic identification, but nonetheless with an identity lower than 97%, 21 were assigned to the Family, Order, Class or Phylum and the remaining couldn't be assigned (Appendix Tab. 5). The 18 matches mentioned before allowed the taxonomic assignment of approximately 87% of the ARMS metabarcoding pre-dereplication total sequences, mainly thanks to the assignment of the first (most abundant) OTU, accounting alone for 78% of the total sequences.

The 241 GenBank release of the MIDORI reference library, after the OBITools pipeline processing detailed in the methods, accounted for 28,008 dereplicated sequences (Appendix Tab. 5). The taxonomic assignment reported 13 ARMS OTUs with a match identity greater than 97%. However, two of those OTUs couldn't be identified at the Species level due to uncertainty in the taxonomic identification, with a successful identification only at the Genus and Family levels respectively. Of the remaining OTUs, 23 were matched at the Species level, with an identity lower than 97%, 37 were assigned to a taxonomic level between the Genus and the Phylum (without counting the two matches mentioned earlier with a high identity percentage), while the remaining were assigned to a higher taxonomic level (Appendix Tab. 5). The first 13 matches allowed the taxonomic assignment of only approximately 7,5% of the ARMS metabarcoding pre-dereplication total sequences, mainly due to the failed taxonomic assignment of the first OTU, at least at the Species level.

The higher number of matches with a sufficient confidence at the Species level performed using the RSRL remarks the importance of creating regional specific libraries with sequences extracted from species typically inhabiting the area investigated (Questel et al., 2021). This is also evident if we consider one of those two matches mentioned earlier, for which the *ecotag* program couldn't confidently identify the belonging species despite the high similarity, as the two sequences that matched the query belonged to two Porifera species, one typically occurring in the Mediterranean Sea, *Phorbas fictitius* (Bowerbank, 1866), and another from New Caledonia, *Hooperia anfractuosa* (Hooper & Lévi, 1993). The other match, instead, was assigned to the genus *Alcyonium* Linnaeus, 1758 (Anthozoa, Cnidaria). However, no correspondent match was found for the same OTU using the RSRL as the resource downloaded from GBIF didn't report any occurrence labeled as "preservedSpecimen" for this genus in the Ross Sea MPA ("GBIF Occurrence Download https://doi.org/10.15468/dl.aq96re"). Nonetheless, as the presence of this genus is known for the area (Smith et al., 2007), a wider download parameter set may be necessary to gather more information

on the area investigated, possibly in combination with more stringent quality filtering of the occurrences retrieved, allowing for the recovery of additional records (*e.g.* "HumanObservation" occurrences). The adopted pipeline allowed the retrieval of records most plausibly associated to voucher specimens, which doesn't increase the overall reliability of the occurrence *per se*, but offer a higher guarantee for replicability, especially considering the application of new technological advances on them (Troudet et al., 2018), and consequently the possibility to update the occurrence information based on the state of the art in biological research.

The amount of confident species matches observed by adopting the two different libraries (MIDORI and RSRL) in the taxonomic assignment resulted opposite to the number of total matches at higher taxonomic levels (Appendix Tab. 5). The sensibly lower number of total sequences in RSRL lowered the odds of a low confidence match, and thus the taxonomic assignment to higher taxonomic levels. However, if we take into account the entire ARMS dataset, the taxonomic assignment didn't show a substantial, overall difference between the application of the two libraries, with a striking similarity in the relative frequency assignment at the Phylum level (Fig. 7a and b). Nonetheless, the wide gap between the known species occurrences and barcodes available for regional-specific libraries may exacerbate this condition, especially considering the high variability in sequence coverage between different taxa (Weigand et al., 2019) as discussed here earlier (Appendix Tab. 4).

The additional sequences extracted from MNA samples allowed the taxonomic assignment of 6 OTUs at the Species level with a confident identity similarity. These OTUs were thus identified as *Lanicides bilobata* (Grube, 1877), *Harmothoe fuligineum* (Baird, 1865) (Annelida), *Myxodoryx hanitschi* (Kirkpatrick, 1907), *Marseniopsis mollis* (E. A. Smith, 1902), *Dendrilla membranosa* (Pallas, 1766) (Porifera) and *Camptoplites bicornis* (Busk, 1884) (Bryozoa), substantially increasing the taxonomic assignment success at the Species level for the entire ARMS dataset (Fig. 7c and d).

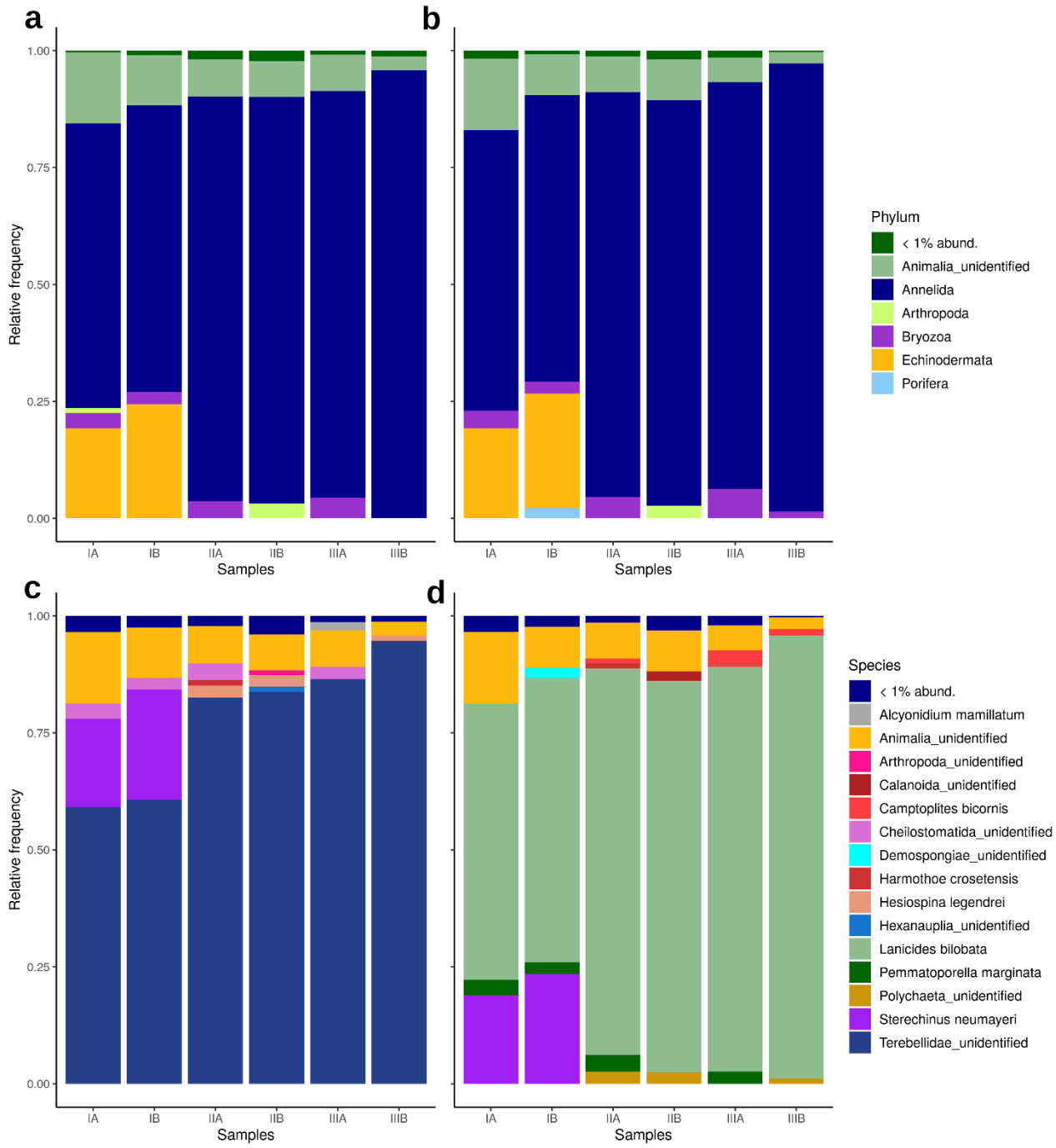


Figure 7: Taxa barplot of the results from the taxonomic assignment using ecotag and the MIDORI reference library (a and c) and the RSRL (b and d). Upper barplots show the results at the Phylum level, whereas lower barplots show the result at the Species level.

One of the advantages of adopting a regional-specific library is the ability to perform taxonomic assignment analyses in a reduced amount of time, due to the lower amount of sequences. The analyses performed for this study were executed on a laptop with 16Gb of memory and an Intel® Core™ i7

processor, and only took 20 minutes approximately for the taxonomic assignment using *ecotag* on the RSRL, whereas approximately 2 hours were necessary for the same analyses performed using the MIDORI library.

All of these reasons suggest the evaluation of a compromise between the adoption of regional-specific libraries, more reliable on the taxonomic assignment at the species level, and a general one, with the advantage of generating a greater number of assignment at higher taxonomic levels.

Chapter 2:

Metabarcoding of the first Autonomous Reef Monitoring Structures (ARMS) deployed in Antarctica (Terra Nova Bay, Ross Sea): molecular insights into three years of growth of benthic communities on artificial substrata.

2.1. Aims of the study:

1. Processing and analysis of the first ARMS deployed in Antarctica with an experimental design that allows the investigation of the development of pioneering communities over a period of one, two and three years since deployment;
2. Comparison of alpha diversity metrics resulting from the application of “DNA metabarcoding” with results obtained from other ARMS deployed in temperate and tropical regions;
3. Analysis of the taxonomic composition and development of the sessile assemblages that colonized the ARMS and qualitative comparison of these results with information previously reported for colonization experiments conducted in Antarctica.

2.2. Introduction

Macrobenthic communities of Terra Nova Bay shallow waters have been extensively studied since the establishment of Mario Zucchelli Station in 1985. The upper sublittoral zone identified between 2 and 3 meters of depth is characterized by a reduced community due to the disturbance of fast ice, ice foot and drifting pack ice (Cattaneo-Vietti et al., 2000), whereas communities found between 12 and 25 meters of depth present high abundance of megaphytobenthos, with Rhodophyta occurring in both erect fleshy forms (*e.g. Iridaea chordata* (Turner) Bory de Saint-Vincent, 1826 and *Phyllophora crispa* (Hudson) P.S.Dixon, 1964) as well as carbonatic crustose forms (*Tethysphytum antarcticum* Sciuto, Moschin & Moro, 2021), usually dominating hard bottom substrata (Gambi et al., 2000; Sciuto et al., 2021). Large sessile metazoans have usually been considered rare or only occasionally observed in this depth range (Cattaneo-Vietti et al., 2000), but members of Porifera, Cnidaria and Annelida have been recently reported as major components of the megabenthic communities in this area between 2 and 20 meters of depth (Kang et al., 2019). A more diverse and abundant megazoobenthic community characterized by sponges, bryozoans, ascidians and anthozoans is observed at greater depths (Cattaneo-Vietti et al., 2000).

Antarctic fouling communities were studied already during first PNRA expeditions, focusing also on those establishing on artificial substrata (Amato, 1990). In the following years, multiple attempts have been performed to retrieve and study these structures, however, only scarce, qualitative information have been obtained on these communities, reporting only the presence of diatoms, bryozoans, serpulid polychaetes and hydrozoans (Cattaneo-Vietti in Bacigalupi and Ramorino, 1994). Since these initial studies dating back to the ‘90s, no new attempts to study fouling communities have been conducted until recently, when new programs on colonization of artificial substrata have been undertaken (Caruso et al. 2018, 2019). The majority of available data about colonization and recruitment of Antarctic benthic communities on artificial substrata has been conducted in other areas of the

continent, especially near Rothera station in the Antarctic peninsula (*e.g.* Bowden et al. 2006), at McMurdo Sound (*e.g.* Dayton 1989) and near Davis Station in the Windmill Islands in east Antarctica (*e.g.* Stark 2008). The main conclusions of these studies reported a general pattern of slow growth in Antarctic communities (Peck, 2018) characterized by sudden and sporadic burst of growth and recruitment often correlated to peculiar oceanographic and physical conditions, such as the duration of the sea-ice cover (*e.g.* Dayton et al., 2016, 2019). However, these studies mainly employed a visual census of the communities inhabiting artificial structures, thus inevitably reducing the analyses to the most abundant and common taxonomic groups (*e.g.* bryozoans and sponges).

High Throughput Sequencing (HTS) techniques come in help by expanding the analyses on the less abundant and more cryptic components of these biological communities, increasing the taxonomic detail but also maintaining a high level of reproducibility (Taberlet et al., 2018). These techniques have been already adapted to the study of fouling organisms for multiple purposes (*e.g.* Azevedo et al., 2020; Zaiko et al., 2016), but less attention has been focused on standardizing the sampling and experimental technology.

Artificial Reef Matrix Structures (ARMS) were first designed by Zimmerman and Martin (2004) in order to provide an artificial structure that would help researchers sample and study the organisms belonging to what was later defined as “cryptobiome”, *i.e.* the community inhabiting hidden spaces of complex 3D environments such as the coral reef matrix (Carvalho et al., 2019). The first ARMS were mainly made of concrete and presented an overly complicated design, useful to attract as much cryptic organisms as possible, but less practical on disassembling and scientific reproducibility. Later on, Leray and Knowlton (2015) devised a simplified version of these structures, now named Artificial Reef Monitoring Structures (ARMS), composed by a specific number of PVC plates stacked on top of each other and presenting alternating layers of crevices open and closed to current flow, still simulating a complex 3D environment. The simplified design of these structures provide an easily

quantifiable sampling methodology without reducing the complexity of the system, allowing at the same time researchers to adopt more advanced technologies to study fouling organisms such as the HTS methodologies mentioned earlier, thanks to the ease of processing PVC plates. Since then, a multitude of programs have been conducted employing ARMS in different areas of the world (https://www.oceanarms.org/), but only very recently different organizations and researcher have worked together to plan and conduct simultaneous monitoring activities at a continental and global level, including Antarctica (Obst et al., 2020), for which, however, no data has ever been published until now. The first regional study on colonization of ARMS conducted at a continental scale encompassing the Mediterranean, Baltic, Black, Red sea and bay of Biscay, was published by Pearman et al. (2020), which revealed a significant relationship between the diversity measurements and environmental descriptors and oceanic distances, further highlighting the usefulness of this methodology in standardized bio-monitoring research.

Changes in growth, recruitment, taxonomic composition, gene expression and intra-specific interactions have been detected in benthic communities on both natural (Barnes, 2013; Barnes et al., 2011; Fillinger et al., 2013; Krzeminska and Kuklinski, 2018) and artificial substrata (*e.g.* Bowden et al., 2006; Dayton et al., 2016), linking these changes with physical (*e.g.* Ashton et al., 2017; Clark et al., 2019; Barnes et al., 2021), oceanographic (*e.g.* Barnes, 2013; Dayton et al., 2016; Dayton et al., 2013) and anthropogenic (Stark, 2008) drivers. All these studies were conducted following experimental procedures that required a minimum of one year (*e.g.* Barnes et al., 2021) of temporal design to a maximum of decades (*e.g.* Dayton, 1989) allowing to identify, or at least suggest, the peculiar environmental conditions that generated these changes. By providing a cost effective and standardized methodology, ARMS may be the best solution for a continuous assessment of the vulnerability of benthic communities to environmental change, especially considering the importance that pioneering communities may have in future seabed assemblages (Barnes et al., 2014). The simple

design of these structures allows to effectively quantify the differences that characterize communities recovered in very different environments, a necessary assumption for bio-monitoring activities in light of environmental changes.

2.3. Materials and Methods

2.3.1. Deployment and recovery of ARMS structures

A set of 6 ARMS in total was deployed by the PNRA SCUBA divers in November of 2015 at 25 meters of depth in the locality of “Zecca”, at the southern entrance to Tethys bay (-74.690°, 164.103°), approximately 500 meters from MZS (Fig. 1).

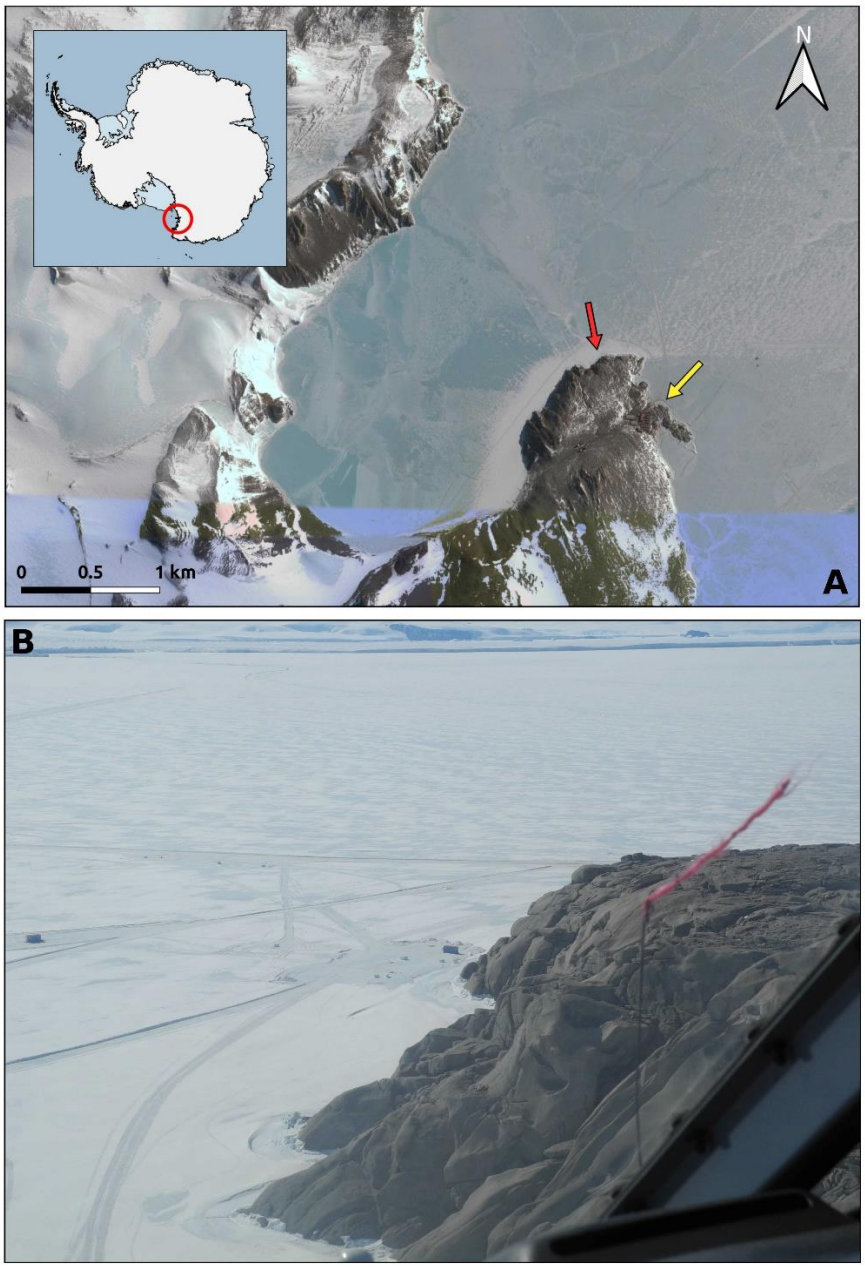


Figure 1: (a) Overview of Tethys Bay in Terra Nova Bay (Red circle in the upper left corner). Yellow arrow indicates the location of MZS, while the red arrow indicates "Zecca" locality, where the ARMS of this study were deployed. (b) Aerial view of "Zecca" locality, taken in November 2018. The "fish-hut" container in the centre of the picture is located near the holes in the ice used by the PNRA SCUBA divers.

The seabed surrounding the ARMS is composed by a heterogeneous, unsorted sediment with both sand, gravel and small cobbles mainly colonized by Corallinales. The area is characterized by a high abundance of *Sterechinus neumayeri* (Meissner, 1900) and *Odontaster validus* Koehler, 1906, which were often found in the site during retrieval of the structures. These structures consist of ten square PVC plates (22.5 x 22.5 x 0.5 cm) stacked on top of each other and separated by 1 cm nylon spacers at the corners of each plate, where four stainless steel bolts are threaded into, holding the entire structure together. This is then fixed on top of a large 45x35 cm baseplate, also in PVC, which allows the entire structure to be anchored to the seafloor by four stainless steel rods passing through large holes in the corners of the baseplate (Fig. 2, https://sketchfab.com/3d-models/monitoring-structures-arms-in-antarctica-b74b246ff5aa420ca8d3e42776e4d3b9). More details on the design and assembly of these structures are provided by the suggested protocols of the Global ARMS Program (https://www.oceanarms.org/).

Figure 2: (a) A new ARMS prior to deployment. (b) One of the ARMS deployed in 2015 and retrieved right after this picture was taken, in November 2018. (c) Downward-facing and (d) upward-facing plates of one of the ARMS retrieved in 2018, after three years since deployment.

The ARMS were retrieved in pairs during November of each of the following three years (2016, 2017 and 2018), thus allowing the recovery of hard bottom benthic communities, with two sampling replicates, after one, two and three years of deployment respectively. These structures were recovered thanks to the help of PNRA SCUBA divers, which covered each retrieved structure with a rigid plastic crate perforated on each side and internally lined with a 100 μm nylon net, in order to avoid the escape of vagile benthic organisms. However, during the recovery of the first pair of structures and one of the second, these crates malfunctioned and thus no analyses on the vagile component of the community inhabiting the ARMS could be performed. The entire recovered structures were then covered in sterile plastic bags and preserved at -20°C until processing.

2.3.2. Processing of the ARMS plate, DNA extraction and sequencing

All structures were disassembled in June of 2018 at the MNA. Each ARMS' plate was carefully removed from the structure, placed in a plastic tray filled with absolute ethanol and photographed with a Nikon D700 equipped with a 105 mm lens. The plate was inspected and, after taking a reference picture, a piece of tissue was carefully sub-sampled from the colony of each morphospecies. This was done until all the most abundant species were sub-sampled. Finally, the benthic organisms growing on the plate were scraped with a trowel and collected in a clean plastic tray and, after all the scrapings from each plate of the structure were collected, blended in a kitchen blender. The homogenized sample was then placed in multiple 50 ml falcon tubes, depending on the amount of homogenized material, for two-thirds of their capacity, filled with absolute ethanol and preserved at -20°C. All the material used in the aforementioned protocol was washed with hydrogen peroxide (35%) and left to dry before processing each new sample.

The falcon tubes were then shipped to the department of Biology and Biotechnologies "Charles Darwin" of the "Sapienza" University of Rome (Italy), where the DNA extraction was performed. The content of one of the tubes for each sample was poured inside a Petri dish covered with sterile

aluminum foil, isolated in a lab oven and left to dry overnight. The dry homogenized material was then mixed and sub-sampled, originating three replicates of 0.25 g approximately, and DNA extraction performed on each extraction replicate using the DNeasy PowerSoil Kit (QIAGEN), following the manufacturer's instructions and thus providing a total of 18 extraction replicates (three DNA extraction replicates for each single structure' homogenize and two structures for each single year). This was adopted after noticing that the total amount of dry homogenized material for the ARMS recovered after one year reached approximately 2 grams, thus hampering the possibility to perform the DNA extraction using kits that require a greater amount of material (*e.g.* the DNeasy PowerMax Soil Kit, QIAGEN) and have been used extensively in studies involving ARMS (*e.g.* Leray and Knowlton, 2015; Pearman et al., 2020, 2016). PCR amplification and sequencing were performed by IGA Technology (Udine, Italy, https://igatechnology.com/). The primers used for the Leray fragment of the COI region (approximately 313bp) were chosen from (Leray et al., 2013) and have the following sequences (Illumina adapters underlined): mlCOIintF - 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGWACWGGWTGAACWGTWTAYCCY CC 3′ and jgHCO2198 - 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTAIACYTCIGGRTGICCRAARAAYCA 3′. The PCR mix consisted in 12.5 µl of 2× KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Woburn MA, USA), 5 µl of each primer and 2.5 µl of microbial DNA at a concentration of 5 ng/µl. The amplification conditions were: 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, followed by a final step at 72 °C for 5 min. A PCR clean-up step was performed using AMPure XP beads (Beckman Coulter) to purify from free primers and primer dimer species. This was followed by an indexing step using the Nextera XT Index (Illumina), to attach dual indices and Illumina sequencing adapters. The PCR program was the same of the amplicon PCR, except for the number of cycles set to 8 instead of 25. Another PCR clean-up step was performed prior to the

quantification, normalization and sequencing using Illumina MiSeq v3 reagents on a 300bp paired end reads MiSeq platform.

2.3.3. Bioinformatic analyses

Two different pipelines were adopted in the bioinformatic analyses. In the first pipeline, only the Antarctic samples were processed, whereas in the second, samples from another study performed outside of the Southern Ocean (Pearman et al. 2020) were included and processed altogether.

2.3.3.1. Analyses on Antarctic samples

Raw COI sequences were quality checked, after demultiplexing, using *FastQC* (Andrews, 2010), and reverse reads of each sample were trimmed of 50 bp at the 3' end to remove the portion of the sequences with the lowest overall quality score. This step was performed with *vsearch* (Rognes et al., 2016), as well as the following merging, allowing a maximum of 10 differences in the alignment and no ambiguous bases. Primers were removed using *Cutadapt* (Martin, 2011) and sequences from all samples were concatenated together and filtered, always using *vsearch*, to remove all sequences with a maximum expected error of 0.25, due to the high overall quality score. *Mothur* (Schloss et al., 2009) was used to remove all the sequences with homopolymers longer than 8 bases, whereas the length filtering (with minimum and maximum length set to 311 and 313 respectively) and dereplication were performed using *vsearch*. After the dereplication, the unoise2 algorithm (Edgar, 2016) implemented within *usearch* (Edgar, 2010), using the command "unoise3", was used to check for chimeras and remove singletons, generating the Zero-radius Operational Taxonomic Units (ZOTUs) fasta file. *Vsearch* was used again for the creation of a count table (command "usearch_global") using a global pairwise alignment with "*id*" equal to 1. The clustering was performed on the ZOTUs using *swarm* (Mahé et al., 2015, 2014) and with "*d*", the clustering distance threshold for the initial phase, set to 13, which has been extensively used for fast evolving markers such as COI (Antich et al., 2021). The taxonomic assignment protocol adopted has been already described in Chapter 1 of this thesis. Briefly,

the seed sequence of each cluster (*i.e.* the representative sequence of each cluster) obtained from the clustering method performed by *swarm* was identified thanks to the *ecotag* program of the OBITools suite (Boyer et al., 2016) and using the Ross Sea DNA barcode library (Chapter 1) as a reference database. The output was transformed to a tabular format using *obitab* and then analyzed by the "Match taxa" tool in WoRMS, to retain the entire taxonomic lineage of each scientific name (more details regarding the taxonomic assignment and following modification on the results are reported in Chapter 1 of this thesis). All of the previously mentioned analyses are reported in the bash script **pipeline_ARMS.sh**. As the count table was created before the clustering, a modified version of the R script *owi_recount_swarm* from the GitHub project "Metabarpark" (https://github.com/metabarpark), here called **recount_swarm.R**, was created in order to aggregate the abundance information of each ZOTU in all the respective clusters. The count and taxonomic assignment tables were then uploaded in R for the following analyses (script **ARMS_barplot.R**).

Taxa barplots were realized at both the phylum and species levels using the R package *phyloseq* (McMurdie and Holmes, 2013), after collapsing together the DNA extraction replicates. Accumulation curves were calculated using *iNEXT* (Hsieh et al., 2016) after collapsing together all the samples corresponding to the same year of recovery, with default settings and on individual-based abundance data. The count table was then stabilized using a variance stabilizing transformation, instead of applying a rarefaction, as suggested by McMurdie and Holmes (2014), using the R package *DESeq2* (Love et al., 2014). Negative values, which in the context of a variance stabilizing transformation indicate that in the original count table those values were more likely to be zero, or in any case negligible, were approximated to 0, as suggested by the *phyloseq* authors (McMurdie and Holmes, 2013) (https://www.bioconductor.org/packages/release/bioc/vignettes/phyloseq/inst/doc/phyloseq-FAQ.html#negative-numbers-in-my-transformed-data-table, last access on October 07/2020). This

approximation allowed the calculation of Bray-Curtis distances for the ordination plots generated through a Non-metric Multidimensional Scaling (NMDS) using *phyloseq*, which was used again for calculating and plotting alpha diversity values of the indices “Shannon” and “Simpson”. Finally, the euler plot was produced using the R package *MicEco* (Russel, 2021).

2.3.3.2. Comparison with ARMS deployed outside of the Southern Ocean

Data from Pearman et al. (2020) were downloaded and integrated with the sequences obtained for this study in order to perform a preliminary comparison between diversity metrics on benthic communities growing on ARMS deployed in temperate or tropical regions with those deployed in a polar region. In order to achieve this, as the ARMS deployment strategy adopted by Pearman substantially differed from that of this study (*i.e.* comprising multiple sites per region and a set of three ARMS per site deployed for approximately one year), only one site per region was randomly chosen from Pearman et al. (2020), and only two ARMS were chosen for that site, in order to more closely resemble the strategy adopted for this study. Moreover, as no DNA extraction replicate was performed by Pearman, only one of the DNA extraction replicates used for this study were chosen for this analysis. If we consider the amount of homogenized sample used for the DNA extraction in the two studies, a great difference is observed. Some biases may have been introduced if we consider these differences, however, as mentioned earlier, the low amount of biological material found growing on the ARMS deployed for this study didn’t allow the use of DNA extraction kits that require 10 grams of tissue, as the one used by Pearman (*i.e.* DNeasy PowerMax Soil Kit, QIAGEN).

The fastq files from the randomly chosen samples (relative to the sessile fraction processing) of Pearman et al. (2020) were downloaded and analyzed together with the chosen extraction replicates from this study. The overall sequence quality was checked with *FastQC* (Andrews, 2010), and reads truncated at the 3’ end keeping 191 and 186 bp for the forward and reverse reads respectively. These length measurements were chosen according to Pearman et al. (2020), which, however, performed

the primer removal before the merging. In this case, as the merging was performed later in the pipeline, the measurements adopted corresponded to those by Pearman et al. (2020), plus the length of the primers (*e.g.* 165+26 bp for the forward reads). This allowed the adoption of a more controlled merging procedure, which was performed using *vsearch* (Rognes et al., 2016) and allowing a minimum overlap length of 10 base pairs with no differences in the alignment and no ambiguous bases in the entire merged sequence. Primers were removed using *Cutadapt* (Martin, 2011) and sequences with a maximum expected error of 1 or more were removed. The dataset was then processed following the same pipeline mentioned in the previous section excluding the taxonomic assignment, which was not considered inherent to the scope of this particular analysis. The final analyses included the estimation of alpha diversity values for the indices “Shannon” and “Simpson” and for the ARMS replicate of each region (and number of years of deployment, considering the structures used in this study) using the R package *phyloseq* (McMurdie and Holmes, 2013). Accumulation curves on individual-based abundance data were created using *iNEXT* (Hsieh et al., 2016) after collapsing all the ARMS replicates for each region.

2.4. Results and Discussion

The structures deployed were gradually colonized over a period of three years, never reaching a complete coverage of the plates' surfaces. Three main assemblages, corresponding to the orientation of the plates' surface, were observed on the structures, with one assemblage mainly composed of terebellid polychaetes localized on the upward surfaces (*e.g.* figures 3 and 4; plates 8,5 and 2 TOP), another one colonizing the downward surfaces (*e.g.* figures 3 and 4; plates 10,8,5 and 2 BOTTOM) and mainly characterized by encrusting and erect bryozoans together with serpulid polychaetes, and another one which colonized the upward surface of the tenth plate on top of the entire structure, almost entirely covering the plate and formed by a diatom biofilm (Figs. 3 and 4; plate 10 TOP).

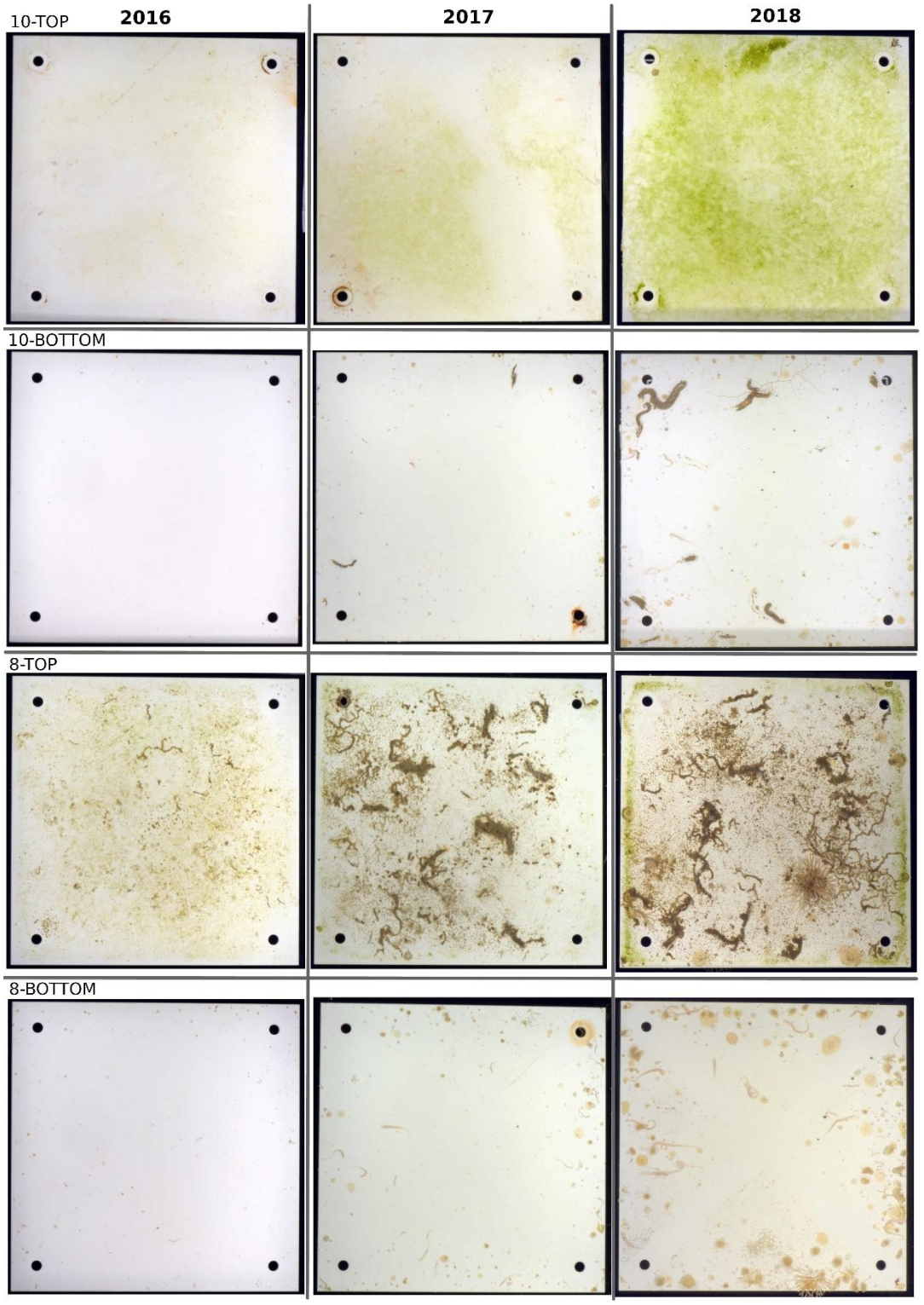


Figure 3: Plates' surfaces prior to scraping for the ARMS sampled after one (2016, left), two (2017, center) and three years (2018, right). Plate's number is showed on the top left corner of each box on the left, indicating also if the surface was upward-facing (TOP) or downward-facing (BOTTOM).

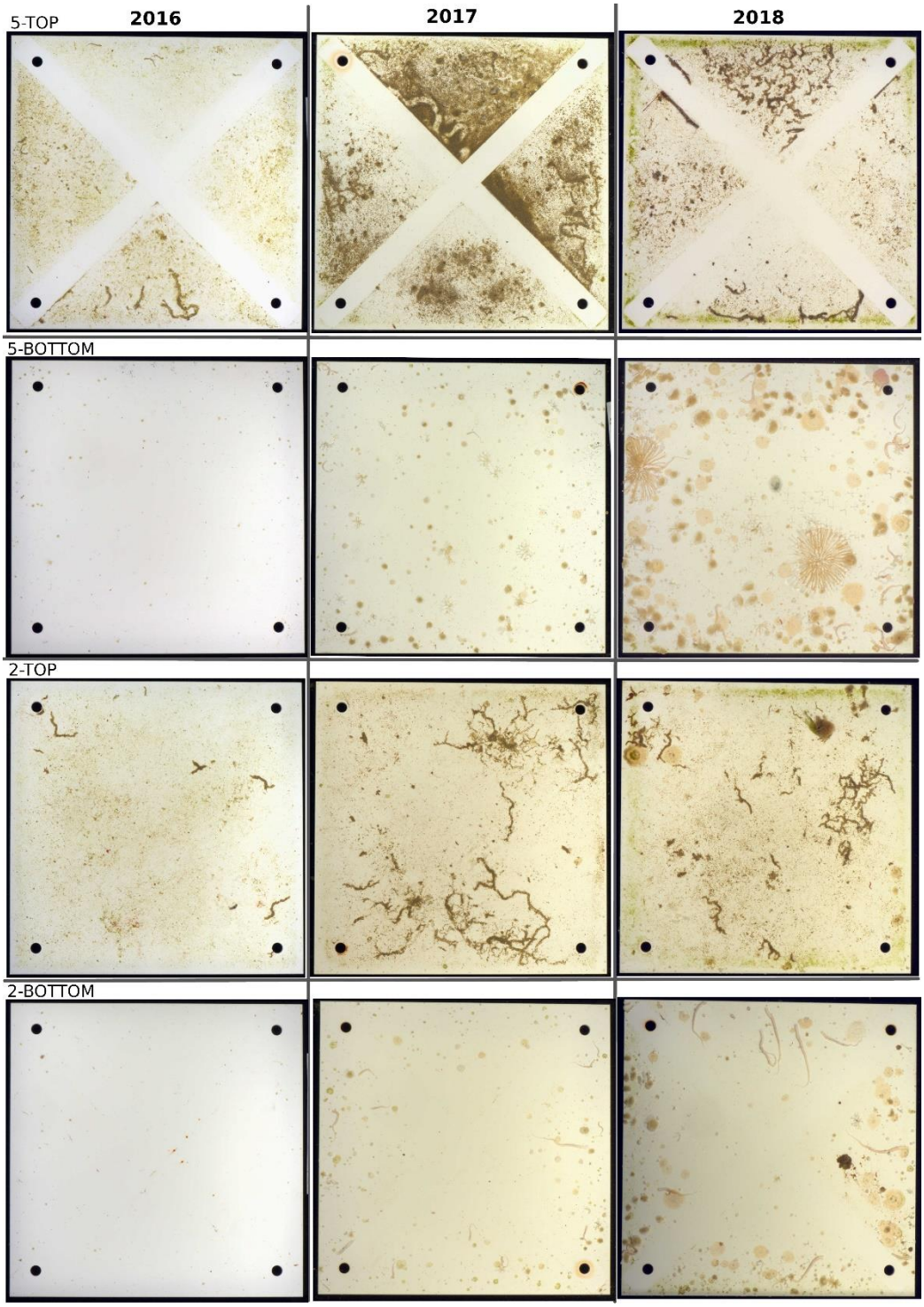


Figure 4: Plates' surfaces prior to scraping for the ARMS sampled after one (2016, left), two (2017, center) and three years (2018, right). Plate's number is showed on the top left corner of each box on the left, indicating also if the surface was upward-facing (TOP) or downward-facing (BOTTOM).

The paucity of colonizing organisms in the ARMS deployed for one year (*e.g.* figs. 3 and 4; year 2016) didn't allow for a general characterization at high taxonomic resolution of the specimens

colonizing the structure, except for the serpulid polychaetes, most certainly belonging to the species *Serpula narconensis* Baird, 1864 (Fig. 5g), and for juveniles of the terebellid species *Lanicides bilobata* (Grube, 1877) (Fig. 5f) whose identification was supported by the taxonomic assignment of the metabarcoding sequences, which will be discussed later. However, multiple small colonies of bryozoans were observed on the downward surfaces and, after examining the colonies that grew on the structures retrieved the following years, were identified as mostly belonging to the genera *Micropora* Gray, 1848 and *Beania* Johnston, 1840. The width of the bryozoan colonies never exceeded 3 millimeters, whereas the terebellids that colonized the upward side of the panels resulted more developed, reaching a maximum of 3 cm in length approximately.

The structures retrieved after two years (Figs. 3 and 4; year 2017) were colonized by a much more recognizable multitude of entrusting and erect bryozoan colonies on the downward surfaces, the most common belonging, as mentioned earlier, to the genera *Micropora* and *Beania* (Fig. 5a and b) for the encrusting colonies, and by some specimens of *Idmidronea* sp. Canu & Bassler, 1920 and *Camptoplites bicornis* (Busk, 1884) (Fig. 5c and d), for the erect ones. A high number of not erect spirorbid polychaetes (Fig. 5h) and only some individuals of the genus *Helicosiphon* Gravier, 1907 (Fig. 5i) were also observed. On the other hand, the upward surfaces were still colonized mainly by terebellid polychaetes, with only some sporadic colonies of *Micropora* and individuals of *Serpula narconensis* Baird, 1864.

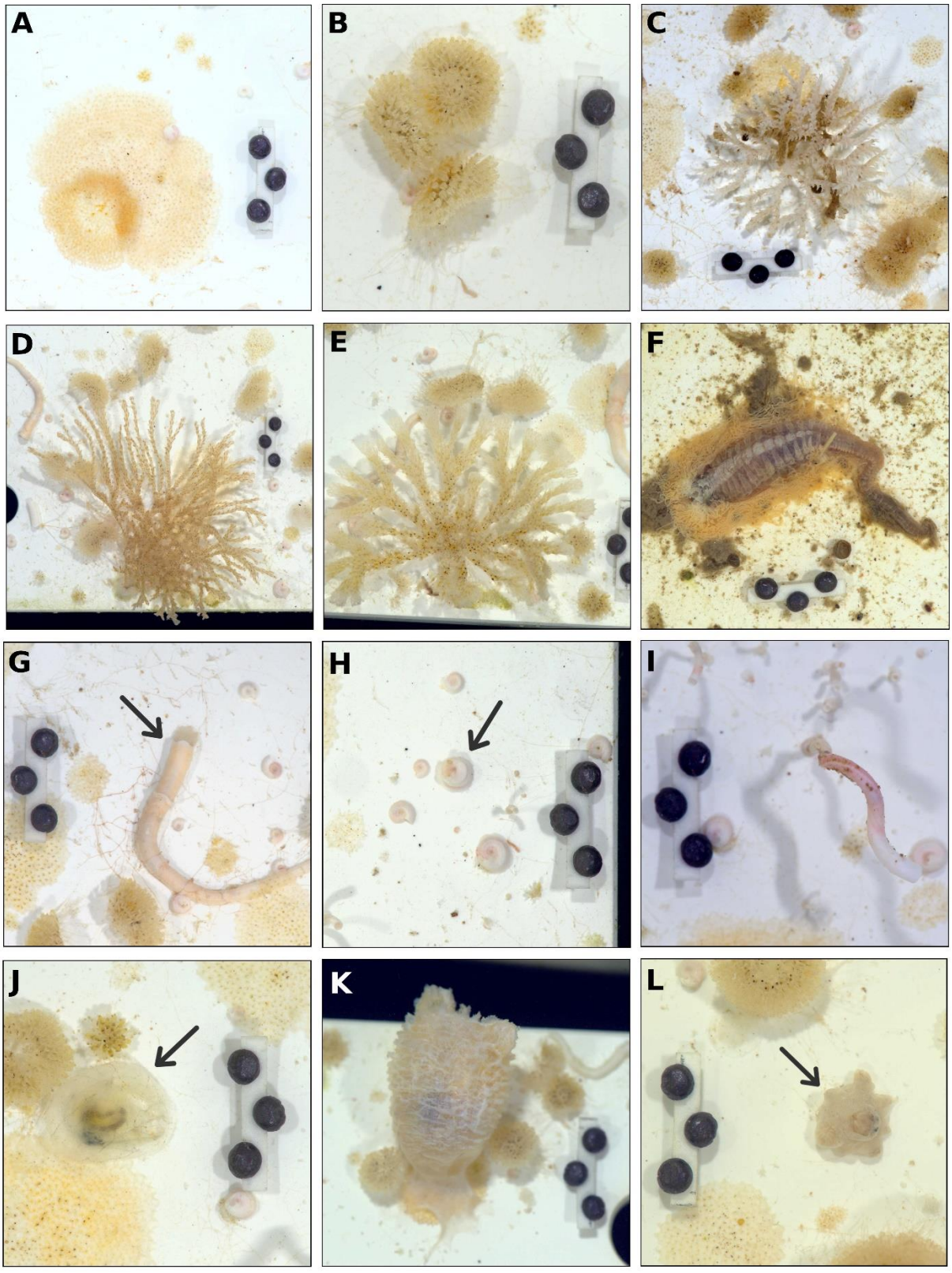


Figure 5: The most abundant taxa of the sessile fraction that colonized the ARMS. Bryozoa; (a) Micropora sp., (b) Beania sp., (c) Idmidronea sp., (d) Camptoplites bicornis (Busk, 1884), (e) Camptoplites tricornis (Waters, 1904). Annelida; (f) Lanicides bilobata (Grube, 1877), (g) Serpula narconensis Baird, 1864, (h) Unidentified serpulidae, (i) Helicosiphon sp. Chordata; (j) Unidentified Ascidiacea, (k) Cnemidocarpa sp. Cnidaria; (l) Alcyonium antarcticum Wright & Studer.

The structures retrieved after three years (Figs. 3 and 4; year 2018) were characterized by a more developed, but nonetheless similar community, in respect to the one observed on the ARMS retrieved after two years. The most striking differences consisted in the presence of some new taxa, however still rarely represented and certainly not dominant, belonging to at least two different species of Ascidiacea (one belonging most certainly to the genus *Cnemidocarpa* Huntsman, 1913, Fig. 5J and k), an anthozoan colony of a few zooids, probably belonging to *Alcyonium antarcticum* Wright & Studer, 1889 (Fig. 5l), some additional and not previously reported bryozoan species belonging to the genus *Camptoplites* Harmer, 1923 (possibly *Camptoplites tricornis* (Waters, 1904); Fig. 5e) and *Disporella*, and one erect hydrozoan colony. The encrusting bryozoan colonies belonging to the genus *Micropora* were significantly more developed in respect to those retrieved in the previous years, with a maximum width exceeding 26 millimeters, whereas one erect colony of the species *Camptoplites bicornis* (Busk, 1884) reached a maximum of 72 millimeters of width. *Helicosiphon* sp. individuals resulted also more developed and abundant in respect to the structures retrieved after two years.

All structures were also colonized by other taxa, such as some erect and branching hydrozoan forms and what appears to be parenchymella larvae of Porifera (Fig. 6a and b), the latter possibly reflecting a photonegative behavior (Maldonado et al. 2003), considering that they were always observed in the downward-facing sides of the plates. These taxa, instead of showing a gradual increase in the number of colonies, individuals or in biomass, were observed in approximately the same quantities on all the structures. The same was observed for some benthic foraminiferans (Fig. 6c), which were often found agglutinated to terebellid tubes, together with some empty valves of juveniles of *Adamussium colbecki* (E. A. Smith, 1902) (Fig. 6d).

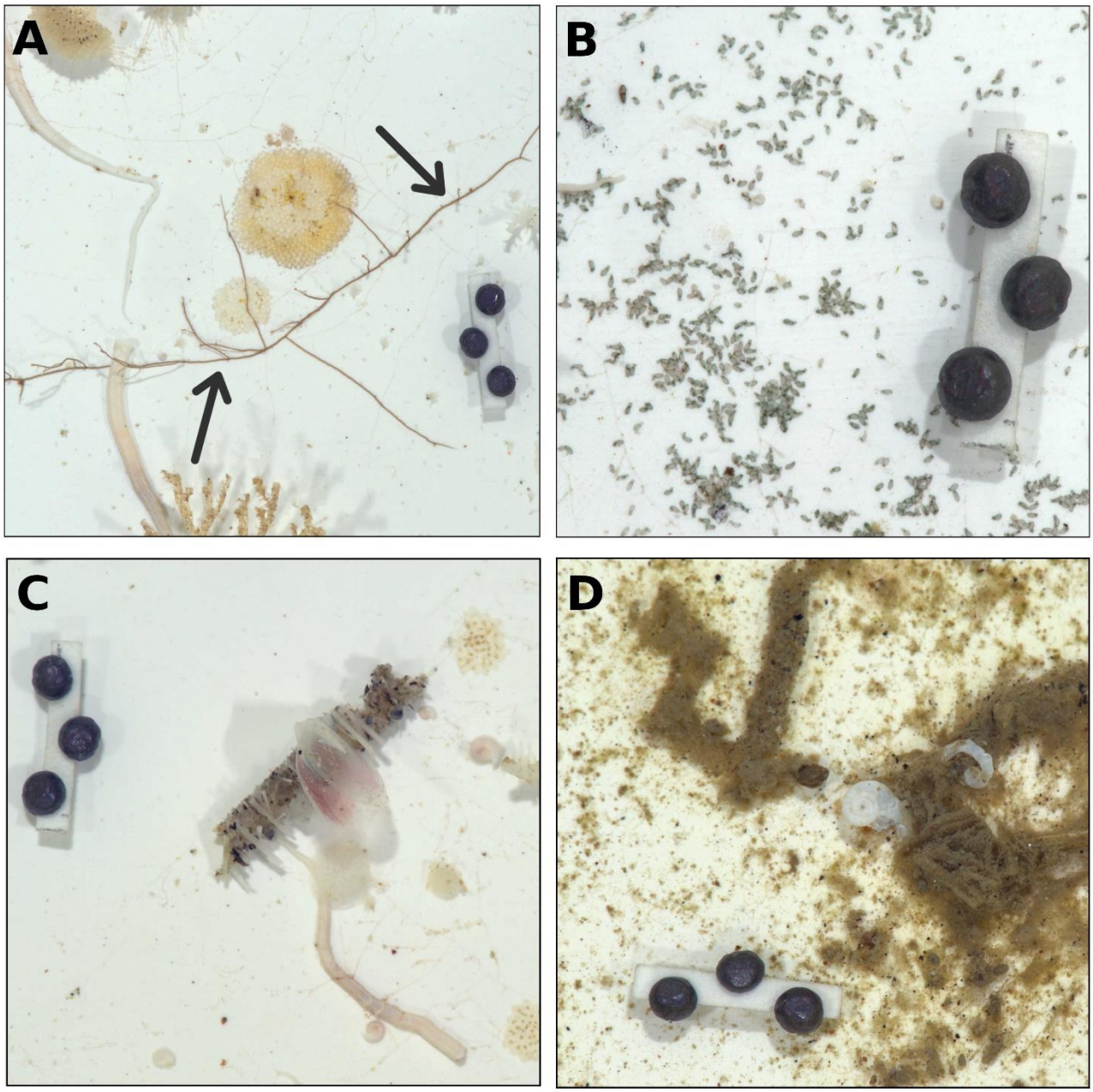


Figure 6: Less abundant taxa found on multiple ARMS. (a) Unidentified hydrozoan colony. (b) Unidentified parenchymella larvae. (c) Juvenile of Adamussium colbecki (E. A. Smith, 1902). (d) Unidentified foraminiferan attached to terebellid tube.

The community observed in the three different years resulted very different from the assemblages found on the seabed surrounding the ARMS deployment location, which, as mentioned earlier, was dominated by well-established populations of macrophytobenthos of crustose and erect forms of Rhodophyta (Cattaneo-Vietti et al., 2000). However, a great resemblance was found with the epiphytic communities commonly found growing on the erect forms of *Phyllophora antarctica*

A.Gepp & E.S.Gepp, 1905, whose role in supporting the diversity of benthic communities has already been highlighted in Terra Nova Bay (Thrush et al., 2006). These epiphytic communities were described as mainly composed by hydroids, serpulids and bryozoans, the latter especially represented by *Beania livingstonei* Hastings, 1943 and *Celleporella antarctica* Moyano & Gordon, 1980 (Cattaneo-Vietti et al., 2000), thus greatly resembling some of the most common taxa found colonizing the structures used for this study (Figs. 5 and 6). The same study reported the frequent recurrence of *Micropora brevissima* Waters, 1904 in micro-communities composed by foraminiferans, other encrusting and calcified bryozoans and by spirorbid polychaetes living on the lower valves of *Adamussium colbecki* (E. A. Smith, 1902) empty shells on soft bottoms between 20-30 and 60-70m of depth. Rosso and Sanfilippo (2000) reported how many of the organisms found on algal forms, between 4 and 35 meters of depth, are encrusting, r-strategist bryozoan species, rapidly reaching the mature stage, and thus important pioneering organisms on ephemeral substrata.

Different Italian researchers reported the presence of bryozoans, serpulid polychaetes and hydrozoans growing on artificial structures after three years of deployment (Cattaneo-Vietti in Bacigalupi and Ramorino, 1994). Other studies focusing on artificial structures used for colonization experiments also mentioned both the under-representation of typical taxa found in the surrounding area and the similarity with the local assemblage of bryozoan and serpulid communities (Bowden, 2005; Stanwell-Smith and Barnes, 1997). Pioneering assemblages (especially the bryozoan component) have shown a strong correlation between the numbers of taxa in the newly recruited communities and the local resident, mature bryozoan assemblages (Kuklinski et al., 2017), thus further highlighting the pioneering role of these kind of assemblages.

The assemblages here observed were also reported in other colonization studies previously performed in Antarctica, showing a considerable level of similarity with this study. The first colonization study conducted in continental Antarctica and employing artificial structures regularly resurveyed over a

period of three years, was conducted by Bowden et al. (2006) at depths similar to this study and using single acrylic plates deployed at Rothera Point and Anchorage Island (west Antarctic Peninsula). Both cheilostome bryozoans and spirorbid polychaetes were the most abundant groups on the colonized downward surfaces, with the former being represented by the highest number of species. Ascidians, sponges and cnidarians were represented by a reduced number of individuals, more developed only after three years, similarly to what was observed in the ARMS used for this study. Bowden's study did not include in the analyses the base plate located below the main artificial structure, thus excluding the assemblages of crevice-like upward-facing surfaces from the analyses. However, the upward surfaces analyzed by Bowden et al. (2006) were more similar to the upward surface of the tenth plate on top of the ARMS structure and, due to the high level of disturbance by grazers, showed an irrelevant level of colonization, never exceeding 10% of surface cover, with only some bryozoans and spirorbid polychaetes surviving (Bowden et al., 2006). A similar community was observed by Stark (2008), which deployed artificial single tiles situated on top of a trough formed from one-half of a PVC pipe, in the area before Casey station (Windmill Island, east Antarctica). The downward-facing surfaces were, again, almost exclusively colonized by bryozoans and spirorbid polychaetes, with rare and more developed sponges, hydroids and ascidians only after three years of colonization (Stark, 2008). Major differences were observed between upward and downward surfaces, with the upward surfaces partially covered by sediment and diatom biofilm only. The same kind community was reported in other, more complex, manipulation experiments (*e.g.* Ashton et al., 2017; Barnes et al., 2021). Bryozoans, serpulid polychaetes and hydrozoans were also observed on artificial, and natural, substrata monitored for long-term studies in McMurdo Sound (see Dayton et al., 2016), however, the peculiarities of the adopted structures, the different monitoring frequency and the fact that those studies mainly focused sponges do not allow for an in-depth comparison of the assemblages.

Bioinformatics analyses conducted exclusively on the samples obtained from the structures deployed in Antarctica for this study (section 2.3.3.1. Analyses on Antarctic samples), produced a total of 342 ZOTUs, which were clustered in 224 OTUs. Final abundance values add up to 1,520,376 sequences, corresponding to ~48% of the total "raw" sequences. Out of the total 224 OTUs, only 35 were assigned to the species level by the *ecotag* program (Tab. 1), out of which, only 18 were assigned with a sufficiently confident level (>97 in the alignment score of the best match in the reference database).

id	best_identity	order_name	family_name	species_name
Zotu1_1197677	1	Terebellida	Terebellidae	Lanicides bilobata
Zotu20_7189	1	Phyllodocida	Polynoidae	Harmothoe crosetensis
Zotu30_2065	1	Valvatida	Odontasteridae	Odontaster validus
Zotu54_1224	1	Dendrochirotida	Cucumariidae	Staurocucumis turqueti
Zotu119_68	1	Phyllodocida	Polynoidae	Barrukia cristata
Zotu139_46	1	Forcipulatida	Asteriidae	Diplasterias brucei
Zotu164_23	1	Calanoida	Calanidae	Ctenocalanus citer
Zotu284_4	1	Spionida	Spionidae	Laonice weddellia
Zotu275_3	1	Dendroceratida	Darwinellidae	Dendrilla membranosa
Zotu332_2	1	Heteronemertea	Cerebratulidae	Parborlasia corrugatus
Zotu39_1030	1	Poecilosclerida	Tedaniidae	Tedania charcoti
Zotu105_107	1	Poecilosclerida	Hymedesmiidae	Myxodoryx hanitschi
Zotu3_101733	0.99	Camarodonta	Echinidae	Sterechinus neumayeri
Zotu21_11141	0.99	Cheilostomatida	Bugulidae	Camptoplites bicornis
Zotu60_1078	0.99	Perciformes	Nototheniidae	Pleuragramma antarctica
Zotu62_607	0.99	Phyllodocida	Polynoidae	Harmothoe fuligineum
Zotu204_13	0.99	Calanoida	Acartiidae	Paralabidocera grandispina
Zotu192_17	0.98	Littorinimorpha	Lamellariidae	Marseniopsis mollis
Zotu57_721	0.96	Alcyonacea	Primnoidae	Onogorgia nodosa
Zotu335_2	0.94	Cheilostomatida	Bugulidae	Camptoplites bicornis
Zotu266_3	0.94	Calanoida	Acartiidae	Paralabidocera grandispina
Zotu141_71	0.93	Phyllodocida	Polynoidae	Harmothoe crosetensis
Zotu267_3	0.9	Poecilosclerida	Hymedesmiidae	Myxodoryx hanitschi
Zotu240_5	0.89	Leptothecata	Obeliidae	Obelia bidentata
Zotu243_6	0.88	Leptothecata	Obeliidae	Obelia bidentata
Zotu286_3	0.82	Terebellida	Terebellidae	Lanicides bilobata
Zotu315_2	0.81	Poecilosclerida	Tedaniidae	Tedania charcoti
Zotu203_15	0.8	Poecilosclerida	Tedaniidae	Tedania charcoti
Zotu65_542	0.8	Cheilostomatida	Sclerodomidae	Systenopora contracta
Zotu85_208	0.8	Cheilostomatida	Sclerodomidae	Systenopora contracta
Zotu106_197	0.8	Cheilostomatida	Sclerodomidae	Systenopora contracta
Zotu6_29149	0.77	Cheilostomatida	Smittinidae	Pemmatoporella marginata
Zotu173_21	0.77	Terebellida	Cirratulidae	Cirratulus cirratus
Zotu12_5656	0.76	Cheilostomatida	Bugulidae	Camptoplites bicornis
Zotu317_2	0.72	Monhysterida	Monhysteridae	Halomonhystera disjuncta

Table 1: Taxonomic assignment at the species level performed by the ecotag program and showing the alignment score of the best match in the reference database (best_identity).

The taxonomic assignment of the 224 OTUs showed a very stable community mainly characterized by the dominance of Annelida, Echinodermata and Bryozoa, followed by Arthropoda and Porifera,

with the major differences apparently discriminating the community found on the structures retrieved after one year from the following ones (Fig. 7a).

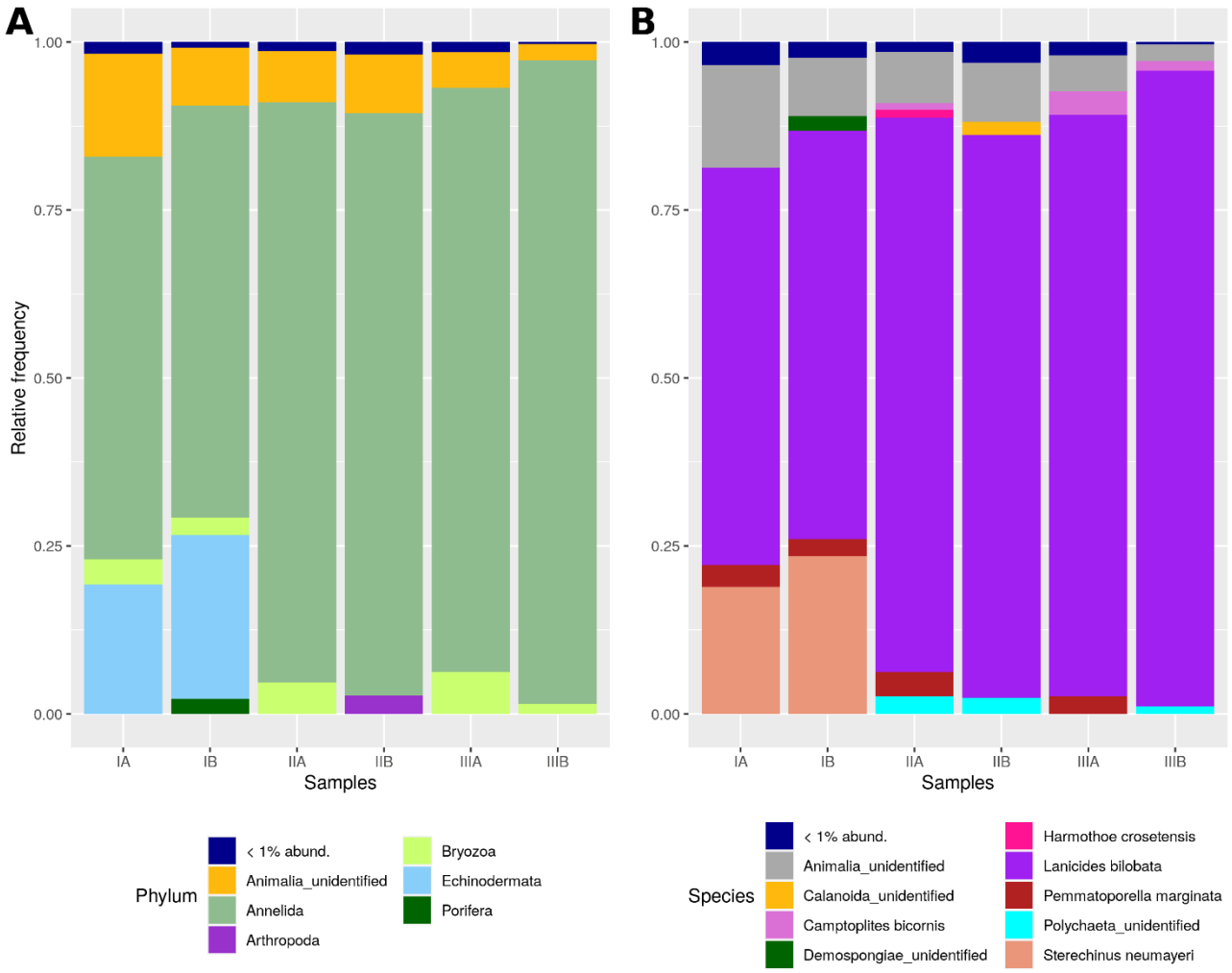


Figure 7: Taxa barplot at the Phylum (a) and Species (b) level, from the taxonomic assignment of the OTUs.

The most striking result is the overwhelming abundance of Annelida sequences in the entire dataset, which characterize all the structures since the first year of deployment (Fig. 7). Moreover, the great majority of these sequences were assigned to the same species, *Lanicides bilobata* (Grube, 1877), apparently identifying this species as the most abundant on all structures (Fig. 7b). This species has already been reported for Terra Nova Bay (Cantone et al., 2000) and was identified as a microphagous detritus feeder of sandy mud (Hans et al., 2011) but also on hard bottoms (Cantone and Di Pietro, 2001), between 10 and 280 meters of depth. Bowden et al. (2006) reported the presence of crevice-

occupying deposit feeding taxa, such as terebellid polychaetes, in the space between the bottom and the artificial structures used for that colonization study, in a very similar way to what was observed in this study. However, the overwhelming proportion of sequences here identified as *L. bilobata* doesn't seem to correspond to the amount of abundance visually detected on the plate's surface, especially considering that, except for some well-developed adults on the ARMS retrieved after three years (Fig. 5f), most of the individuals found on the structures were juveniles rarely exceeding a few millimeters in thickness (*e.g.* figs. 3 and 4; plates BOTTOM), differently from the bryozoans, which established numerous colonies on the ARMS since the first year.

The most probable explanation for this discordance may be identified in the processing of the homogenized sessile fraction which, differently from the standard procedure adopted and suggested by the Global ARMS Program (Smithsonian Institution; https://www.oceanarms.org/protocols/processing/plate-scrapings), due to the particular condition in which the ARMS were preserved before processing, could not be rinsed with filtered seawater, dried and mixed as a whole before being divided in multiple subsamples and preserved again in a chosen medium. Despite the procedure adopted here required the mixing of each dried subsample content as a whole before further subsampling the amount required by the DNA extraction protocol, as not all the original material was dried and mixed a stratification of the homogenized sample may have favored the subsampling of homogenized material from soft-bodied organisms (in this case mostly represented by *L. bilobata*) in contrast to calcified and heavier organism (in this case bryozoans and serpulids).

Moreover, by looking at the taxonomic assignment results at the species level (Fig. 7b), it appears evident that the contribution of some of these groups may be the result of the amplification of non-sessile taxa, such as the *Sterechinus neumayeri* (Meissner, 1900) or the OTUs identified as Calanoida, which certainly belong to the vagile fraction colonizing the ARMS and have not been successfully

removed from the sessile fraction. Considering the already mentioned low amount of biomass of the entire sessile fraction of the ARMS retrieved after one year (< 2g), the high sensibility of the DNA extraction may have favored the detection of taxa not belonging to the sessile fraction (most probably from body parts of vagile organisms such as pedicellariae of echinoids), especially considering the virtual absence of the same proportion of *S. neumayeri* sequences in the structures retrieved the following years, despite many individuals were observed grazing on all structures (Fig. 2).

The adoption of the protocol suggested by the Global ARMS Program has been already identified as the preferable choice, especially in the light of standardization, but also thanks to its ability to more closely resemble other, more traditional analyses such as statistical sessile organisms' coverage from images (Ransome et al., 2017). However, adopting this preferred protocol in remote areas such as Antarctica is nonetheless harder to perform, as time, resources and personnel are a limiting factor and long-term programs involving the use of ARMS in colonization studies in Antarctica must address this issue beforehand. Nonetheless, the taxonomic assignment performed by the "DNA metabarcoding" analyses correctly identified the same most abundant taxa that were observed on the structures, being composed mostly by members of the Annelida and Bryozoa phyla (Fig. 7a).

All samples from the different structures were clearly discriminated based on the corresponding years of retrieval in the NMDS (Fig 8a). Different taxonomic groups appear to characterize the time elapsed since deployment, with Annelida, Arthropoda and Porifera contributing to all the different years (Fig. 9). However, the high number of Echinodermata and Arthropoda sequences, with the latter group exclusively identified as originating from copepods, as mentioned earlier most probably belong to whole vagile organisms (or body parts) that were not successfully separated from the sessile fraction. The only OTU belonging to Chordata was identified at the species level and resulted belonging to *Pleuragramma antarctica* Boulenger, 1902, probably originating from eggs' remnants descending from the above sea-ice (Vacchi et al., 2012), again possibly reflecting the aforementioned effect. On

the other hand, other taxonomic groups which typically include sessile species, such as Cnidaria and Mollusca, seem to contribute mostly to the community developed after two and three years. Surprisingly, no OTU was identified as Ascidiacea, despite some well-developed individuals were observed on the plates' surfaces.

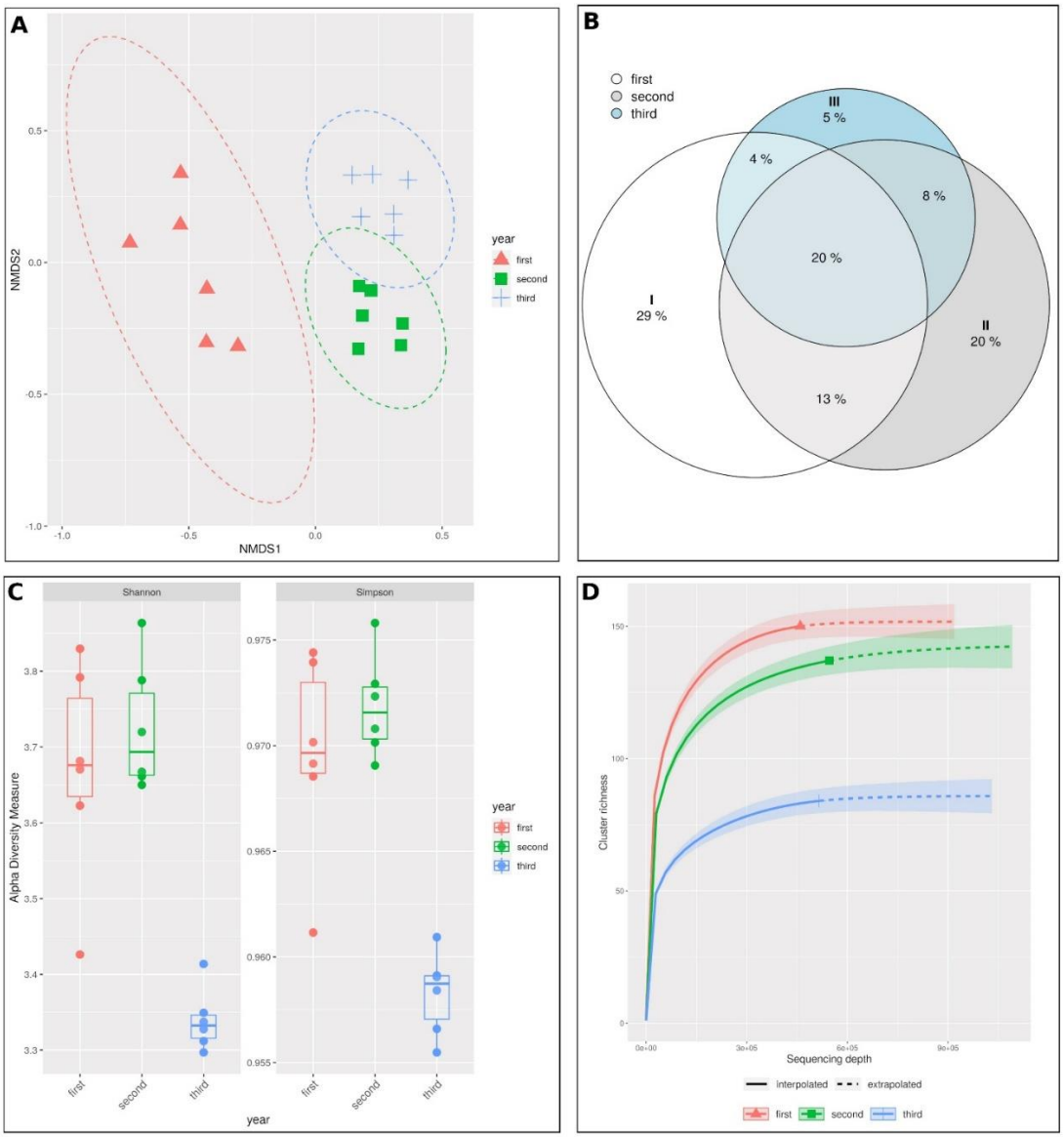


Figure 8: (a) NMDS of all the DNA extraction and year replicates of the ARMS. (b) Euler plot showing the OTUs shared between samples from different years of colonization. The subset area is proportional to the percentage of OTUs shared respective to the total number of clusters in the dataset. (c) Boxplots of Shannon and Simpson metrics for each year of colonization. (d) Accumulation curves for each year of colonization.

However, as mentioned in the first Chapter of this thesis, barcode completeness is still a major limiting factor in metabarcoding studies, and Ascidiacea are one of the groups with the lowest

sequence coverage in the Ross Sea, with only 30% of the species represented by a barcode sequence (Appendix Tab. 4). The same conclusion can be drawn for bryozoans, for which, despite the presence of many sequences identified at least at the phylum level (Fig. 7a), only 4% of the species are represented by a COI sequence in the reference databases (Appendix Tab. 4). For this reason, the apparent exclusive contribution of bryozoan sequences to only the ARMS retrieved after two and three years (Fig. 9) should be considered in the light of which bryozoan species were actually identified by the taxonomic assignment, as it could not successfully identify the most common bryozoan species (belonging to *Micropora* and *Beania*) observed on all the structures since 2016.

Nonetheless, as mentioned earlier, the protocols suggested by the Global Arms Program have been designed to account for, and reduce the impact of, such issues. Many different suggested procedures, such as the barcoding on the major taxa found on the structures (see first paragraph in section 2.3.2. “Processing of the ARMS plate, DNA extraction and sequencing”), could not be performed yet for this study, but have been proposed to increase the number of taxonomic assignments in light of the renown limited reference library completeness (see Leray and Knowlton, 2015). Consequently, by performing such procedures, the majority of unidentified sequences would be properly resolved by the taxonomic assignment, and in this case improve the different contribution of each taxonomic group to the differenciation of the ARMS deployed for a different amount of time (Fig. 9), as well as reduce the apparent overwhelming abundance of terebellid sequences (Fig. 7b).

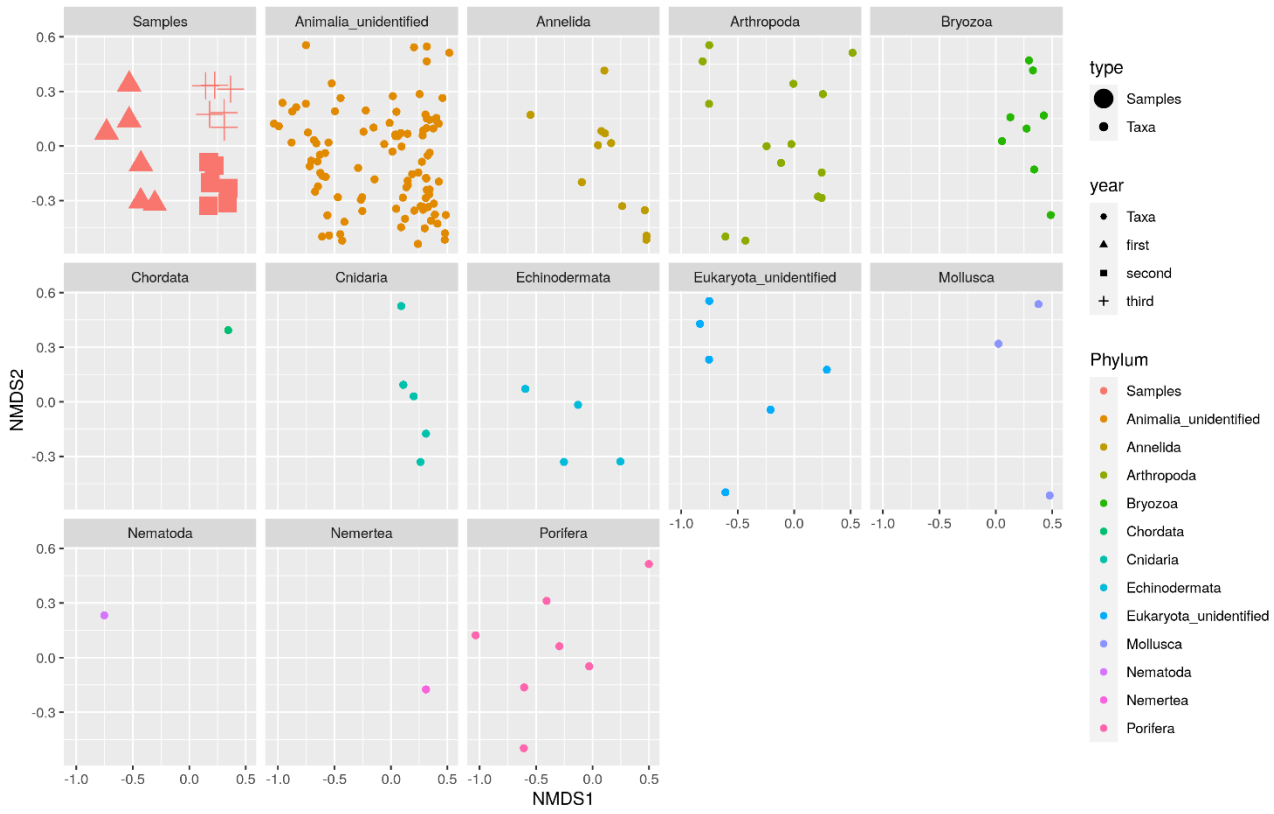


Figure 9: NMDS of both the "Samples" in the upper left corner (as the one in Figure 8a) and of the OTUs belonging to all the different Phyla observed.

The gradually increasing, but nonetheless never complete, colonization observed over a period of three years, supports the interpretation of these communities as characterized by a slow development and a reduced diversity. In fact, only a small proportion of the OTUs were exclusively retrieved from the structures deployed for three years, meaning that each year only a reduced number of new species apparently settled on the ARMS (Fig. 8b). Despite the clear discrimination between the communities at the different years of retrieval revealed by the NMDS (Fig. 8a), an important portion of the OTUs (20%) was shared between each year and 45% between at least two years (Fig. 8b), suggesting that the overall composition, at least regarding the most abundant taxa, after three years of colonization was not generally different from the previous years. This condition reflects a loss in diversity from the second to the third year (Fig. 8c) and a well-defined community since the first year of colonization (Fig. 8d).

These results apparently conflict with the visual account of species on the plates' surfaces, which revealed an apparent increase of the number of species, especially in the ARMS retrieved after three years. However, the reduction of species richness between the first and third year of colonization has been already reported for other locations at a depth similar to this study (*e.g.* Anchorage Island in Bowden et al., 2006), most probably due to post-settlement processes strongly affecting sessile assemblages, with crevice-occupying deposit feeding taxa directly influencing the community, together with a range of grazers and predators (Bowden et al., 2006). In McMurdo Sound, higher recruitment was detected on artificial structures monitored in long-term colonization studies rather than on natural substrata, an effect of larval filter determined by well-established epifaunal communities (Dayton et al., 2016; Kim et al., 2019). In this context, the increased number of taxa visually observed on the plates' surfaces in the structures retrieved after three years should necessarily be considered keeping in mind the reduced dimensions, and thus detectability, of larvae or small colonies, which were most probably already present on the structures since the first or at least second year of deployment.

The bioinformatic analyses performed on the samples from this study and Pearman et al. (2020) (see section 2.3.3.2. "Comparison with ARMS deployed outside of the Southern Ocean") further highlighted the reduced diversity of the community that colonized the ARMS deployed in Antarctica. In fact, Shannon and Simpson metrics were significantly lower for Antarctic samples in respect to those obtained from ARMS deployed in temperate and tropical regions (Fig. 10). However, these results greatly contrast with Pearman et al. (2020) interpretation of the environmental conditions that influenced the diversity between the different regions, as they reported a particular negative association between OTU richness and local Sea Surface Temperature range (SSTr). In fact, compared to the SSTr reported in Pearman et al. (2020), the SSTr usually observed in Antarctica, at least for shallow waters below 20 meters of depth, is extremely low (Barnes et al, 2006), indicating a condition

of environmental stability that, differently from Pearman's results did not associate with a higher diversity. The diversity values here reported (Fig. 10) also showed that, regardless of the number of years of colonization, pioneering communities in Antarctica are considerably less diverse than those colonizing artificial structures in temperate and tropical regions.

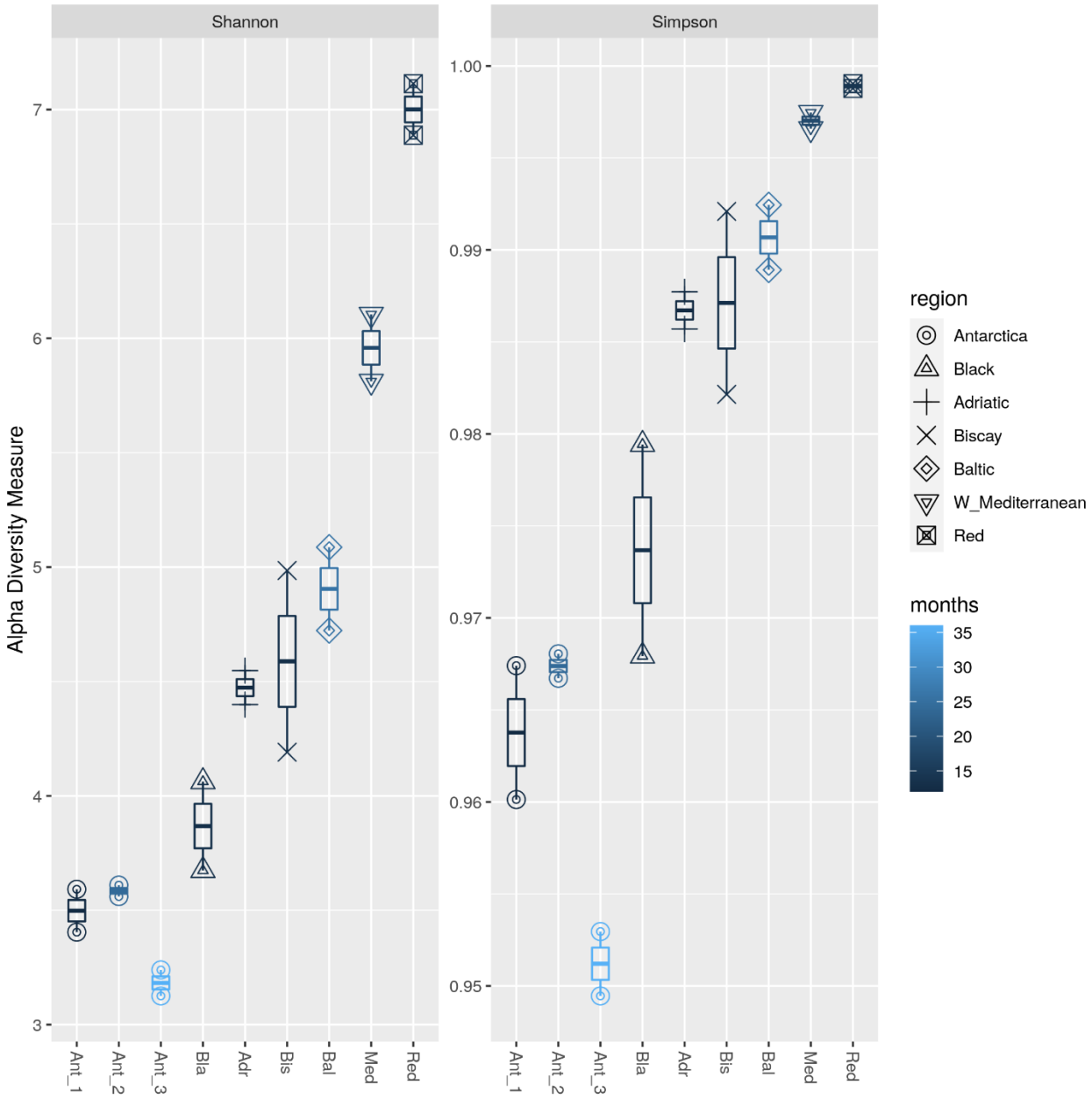


Figure 10: Boxplots of Shannon and Simpson metrics for each Antarctic and Pearman et al. (2020) sample used in the comparison between polar and non-polar areas. Continuous blue bar ranges from 12 to 36 and indicates the color corresponding to the number of months each ARMS was deployed for.

Antarctic benthic communities are considered more diverse than previously thought, exhibiting a complex, functional diversity more comparable with temperate and tropical regions in many cases (Clarke, 2008; Gili et al. 2001), especially for specific taxonomic groups (Chown et al. 2015). In this context, the presence of a lower number of taxa (or OTUs) after three years of colonization in the Antarctic ARMS compared to the same kind of structures deployed in temperate and tropical regions, but for only a third of the time, would nonetheless inevitably identify the Antarctic pioneering communities as characterized by a much slower development of the community. Barnes and Conlan (2012) reported that long-term artificial substrata colonization studies undertaken for 7 years have shown a gradual build-up of diverse fauna, nonetheless characterized predominantly by pioneering species. Sponges, which are commonly represented in epibiotic communities on the seabed before Davis station (Stark et al. 2016), were virtually absent after 1 year of deployment of artificial substrata and well established only after 9 years, while spirorbid polychaetes and bryozoans were present at every sampling interval (Stark 2008 and Clark et al. 2011). A similar pattern was observed in the ARMS used for this study if we consider some relatively large ascidiacean individuals whose presence could be revealed only in the ARMS retrieved after three years (Fig. 5j and k).

Antarctic sessile assemblages might take even 3 times longer to reach 50% of substratum coverage compared to assemblage at temperate latitudes (Bowden et al., 2006). A subset of the ARMS processed by Pearman et al. (2020) was previously published in David et al. (2019), which focused on the visual census of the community colonizing the structures and reported, after only 12 to 16 months of deployment, from a minimum of 50 to a maximum of 75% of the plate's surfaces covered by sessile organisms, not reflecting the colonization observed on the Antarctic ARMS even after three years, and corroborating Bowden's conclusions. Accumulation curves further remarks the distance between polar and non-polar ARMS, showing that, even after reaching the plateau, the number of

OTUs detected in sessile communities on polar ARMS were consistently lower than those found on other regions, for both the colonizing time of one year (Fig. 11a) and more than one year (Fig. 11b).

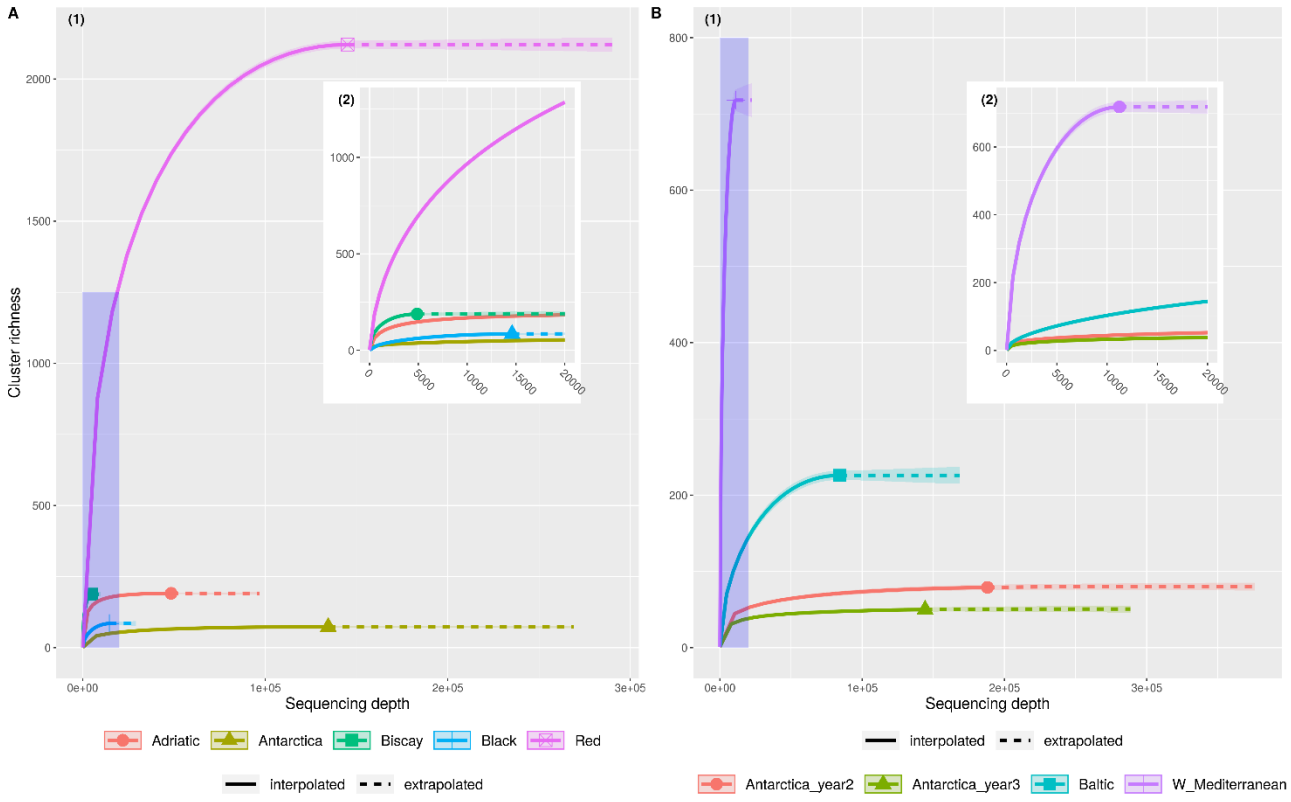


Figure 11 – Accumulation curves for the Antarctic and Pearman et al. (2020) samples deployed for 12 months (a) or more (b). Number in brackets indicate the entire plot (1) and the same analysis run with a maximum limit of 20,000 sequences (2) corresponding to the blue shaded area of the entire plot.

Temperature has taken attention in old literature as the main driver of slow growth in polar regions (Peck, 2018), however, this interpretation has been reconsidered since the first long-term studies on growth of the Antarctic benthos (Dayton, 1989). Barnes (2013), by commenting of Fillinger et al. paper (2013) which revealed an increase in abundance and biomass of Hexactinellida in the region freed from the collapsing Larsen Ice Shelf, reported how different components of Antarctic benthos, like primary consumers (*e.g.* bryozoans) show a substantial increase in growth corresponding to a lower persistence of sea ice cover, thus further reconsidering the supposed major influence of temperature on growth. The effects of changes in sea ice cover were not only recorded for well-established populations (Cummings et al., 2018), but also on recruitment (Dayton et al., 2016),

especially on artificial rather than natural, already colonized substrata (Kim et al., 2019). The complexity of environmental changes in polar regions, due to the interaction of multiple physical properties, prevent us to draw definitive conclusions, especially considering that the communities retrieved on artificial substrata tend to be young and mostly limited to pioneering organisms, limiting our ability to fully relate changes in fouling communities to changes in more developed natural substrata (Barnes et al., 2021).

Nonetheless, the community retained by the ARMS may reflect a “normal” condition, with a slow and gradual growth over a period of three years. However, the lack of similar studies conducted in this region, and thus of a baseline on which to directly compare the results obtained, inevitably hamper the ability to draw meaningful conclusions on the conditions that may have mostly characterized the colonization over the used structures. The results presented here also suggest that Antarctic pioneering sessile assemblages may require decades to show diversity levels comparable to tropical and temperate regions’ assemblages, and that these differences can be quantified using reproducible techniques.

Chapter 3:

Antarctic coastal nanoplankton dynamics revealed by metabarcoding of desalination plant filters: Detection of short-term events and implications for routine monitoring.

3.1. Aims of the study:

1. Look for a correspondence between levels of particulate matter in the seawater and the filter replacement rate;
2. Explore the composition and short-term dynamics of the nanoeukaryotic and particle-attached bacterioplankton communities collected by 5 μm mesh cartridge filters during the Antarctic summer in 2012 and 2013;
3. Address some of the potential issues on the sampling and extraction protocol with the final, future aim of achieving a standardized protocol to be applied on a more general scale.

3.2. Introduction

In the last decades, fine-scale studies on plankton diversity have acquired an increasing importance and attention (Moreira and López García, 2019). Notwithstanding the fact that we are aware that major changes are affecting oceans' functioning, we still lack an effective and internationally coordinated strategy to better detect the effects of these changes (Bindoff et al., 2019). The biggest obstacles are due to the intrinsic variability of spatial and temporal plankton dynamics, coupled with a plethora of methodologies available for plankton biodiversity monitoring. These two aspects exert a synergistic negative effect, overall causing a limited effectiveness in our capability to draw meaningful conclusions on ocean ecosystems state and evolution (Buttigieg et al., 2018; Navarro et al., 2017). This is even more exacerbated in the case of studies of the Antarctic plankton, which is characterized by an intrinsic extreme dynamism, with composition and vertical carbon export changing in a matter of weeks to days (Bathmann et al., 1991; DiTullio et al., 2000; Smith Jr et al., 2003) or even hours, with variations between daytime and night (Celussi et al., 2009). Moreover, a variety of other local, stochastic factors may further sustain this high dynamism, such as water column instability driven by strong winds, that may even suppress the development of phytoplanktic blooms (Moline and Prezelin, 1996) or, in the opposite case, the stratification of the water column in a time frame of days or even hours due to absence of wind-induced mixing (Brandini, 1993). Also coastal pack-ice dynamics can introduce further local variability by moving the location of the sea ice marginal zone and hence the seeding of phytoplanktic blooms (Mangoni et al., 2009), with effects varying at the regional spatial scale and at the seasonal time scale (Dayton et al., 2013). The availability of high resolution time series for Antarctic plankton is thus a crucial point and, at the same time, one of the most difficult research and monitoring tasks, always requiring a great effort to be achieved.

A possible solution or improvement, for achieving high-resolution time series of Antarctic coastal plankton, could be the analysis of samples automatically collected by research base desalination plants. These facilities were already used in a number of ecological studies as an auxiliary sampling methodology for the collection of additional planktic samples, for the investigation of seasonal variations in the phytoplankton, bacteria and picoplankton (Balzano et al., 2015), for the monitoring of harmful algal blooms in the proximity of desalination plants (Villacorte et al., 2015), or to collect invertebrate larvae (Heimeier et al., 2010a, 2010b). Desalination plants are employed wherever freshwater availability is limited and rely on the use of different pre-treatment filters that intercept water-carried particles and organisms and prevent system clogging, before the final reverse osmosis process (Veerapaneni et al., 2007; Wolf et al., 2005). Regardless of the possible technical differences existing in different desalination plants, all these systems employ filters (usually in form of “bags” and “cartridges”) with decreasing mesh sizes, which are replaced whenever the pressure inside the filter housing increases, *i.e.* when they start to clog. Since the freshwater is constantly needed by research base activities, desalination plants operate continuously, drawing seawater throughout the entire research base opening season, hence representing a potential source of planktic samples constantly collected.

The earliest Antarctic studies of desalination filters were authored about ten years ago by Sewell and Jury (2009, 2011) and were done at the New Zealand’s “Scott Base” (McMurdo Sound, Ross Sea). In these studies, desalination plant “primary filters” (100 µm mesh size) successfully collected representative samples of zooplankton (even without damaging the most delicate larval forms) and disclosed the year-round temporal dynamics of the Antarctic meroplankton (Sewell and Jury, 2009, 2011). These studies were also supported by a qualitative comparison with standard net tows samples collected during the same days in the vicinity of the base, revealing a similar composition between desalination plant filter samples and a more traditional sampling strategy (Sewell et al., 2006; Sewell

and Jury, 2009). Sewell and Jury (2009) recognized the many advantages observed by the application of this method, from the opportunity of sampling regardless of weather and sea ice conditions, to the large amount of seawater filtered by the desalination plant (Sewell and Jury, 2009, 2011). The high filtered-water quantity also enabled the collection of rare species that could have been overlooked by using standard plankton net sampling (Sewell et al., 2006; Sewell and Jury, 2009, 2011).

In this study, the usefulness of samples obtained from the desalination plant filters of a research base ("Mario Zucchelli" station, Terra Nova Bay, Ross Sea) combined with highly reproducible molecular metabarcoding analysis (which further reduce sample processing time, increase data precision and expand the study target to smaller ranges of planktic organisms' sizes) adopted to disclose possible changes in the composition of nanoplanktic communities will be evaluated.

3.3. Materials and Methods

3.3.1. Timeframe of the study and of the considered data sets

The first two objectives of this study employed analyses with a different timeframe. In the first case, all the available satellite data, as well as data from the electronic logbook of the desalination plant, from October to February of 2002 to 2019, were included in the analyses, with the only exception of the 24th expedition (2008/2009) for which no data were available. In the second case, sampling of the desalination plant filters was carried out in January and February of 2012 and 2013 corresponding to a total activity period of the filters examined spanning from the 25th of January to the 4th of February of 2012 and from the 8th to the 25th of January of 2013. Automatic Weather Station (AWS) hourly data were downloaded from mid-October to the end of February for both 2012 and 2013, but only those corresponding to the same timeframe of the sampled filters activity time were used. Satellite data, AWS and 5 µm filter activity time (from the desalination plant logbook) for the entire research base opening season (mid-October to end of February) of the 2011-2012 and 2012-2013 Italian Antarctic Expeditions (XXVII and XXVIII) are showed in the supplementary material (Appendix Fig. 1 and 2).

3.3.2. Description of the desalination facility

The desalination plant of the Italian research station “MZS” is partly located inside the construction area of the station (Figs. 1 and 2) and is composed of different pre-filtration steps, leading to the main and final filtration operated by ceramic filters (Fig. 3). Since MZS operates only during the Austral summer, the desalination plant is closed each year at the end of the expedition (around middle of February) by pumping air in all pipes and valves in order to prevent freezing during the Antarctic winter. At the beginning of each season (around mid-October), pipes are therefore fully clean, with no remaining water from the previous season. The entire MZS desalination plant processes 3.5-4 m^3/h on average. Only part of this water enters the true desalination pipeline where the filters operate.

Given the total volume of the pipeline from the intake to the filters (0.24 m^3) it is possible to estimate that this water mass is replaced approximately 15 times per hour, *i.e.* once every 4 minutes.

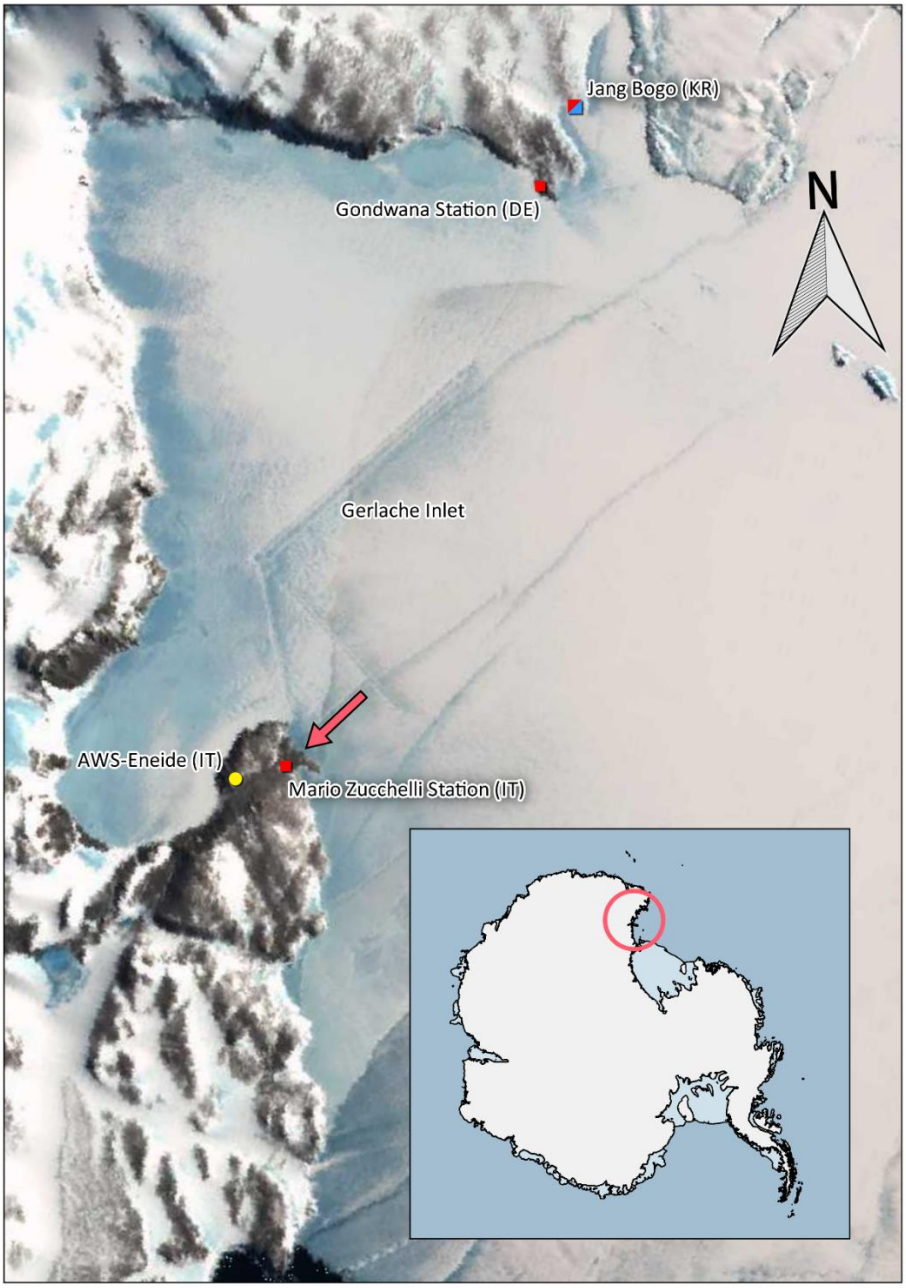


Figure 1: Overview on Gerlache Inlet (Terra Nova Bay, TNB) showing the three research stations operating in TNB: Mario Zucchelli Station (IT=Italy), Gondwana Station (DE=Germany) and Jang Bogo Station (KR=Republic of Korea). The red squares indicate the research stations operating only during the summer, whereas the red and blue square indicate the only all year-round operating research station (Jang Bogo). The map was produced using the collection of datasets "Quantarctica" (Matsuoka et al., 2018) and the 2.18 version of QGIS (QGIS Development Team, 2020). The map depicts the coastline orientation before the desalination plant seawater intake pipe (red arrow) in the locality of Punta Stocchino and of the Automatic Weather Station (AWS) "Eneide" (yellow circle).

Figure 2: Desalination plant of Mario Zucchelli station. (a) View of the plant pump shed in the locality of Punta Stocchino. (b) 25 µm (left) and 5 µm (right) filter housings in the desalination plant powerhouse. (c) New cartridge filters (5 µm) just replaced before the closure of the lid of the filter housing. From Cecchetto et al. (2021).

The desalination facility starts with the seawater intake pipe (-74.6936°, 164.1185°), opening at a depth of 4 meters in the locality of “Punta Stocchino” (Fig. 1). From there, a series of pipes (diameter of 2 inches) and valves allow the water to flow directly to the main powerhouse, distant approximately 120 meters from the intake pump shed. Here the first steps of filtration are obtained through a filter packed with anthracite, followed by polyester bag filters of 25 µm mesh size, a heat exchanger (which brings the seawater temperature to 10° C to maximize the efficiency of the final ceramic filters) and a final set of filters made by polypropylene cartridges of 5 µm mesh size, which were the focus of the present analysis.

The electronic logbook of the desalination plant was inspected to gather all the available historical records for cartridge and bag filters activity and replacement, as well as the amount of consumed water at the research base. All the timings for the filter replacements, together with the activation and turn-off of the desalination plant for technical purposes, are recorded in the logbook. Thus, it is possible to obtain the exact number of hours each filter has been filtering before its replacement, done in order to avoid reaching the clogging limit.

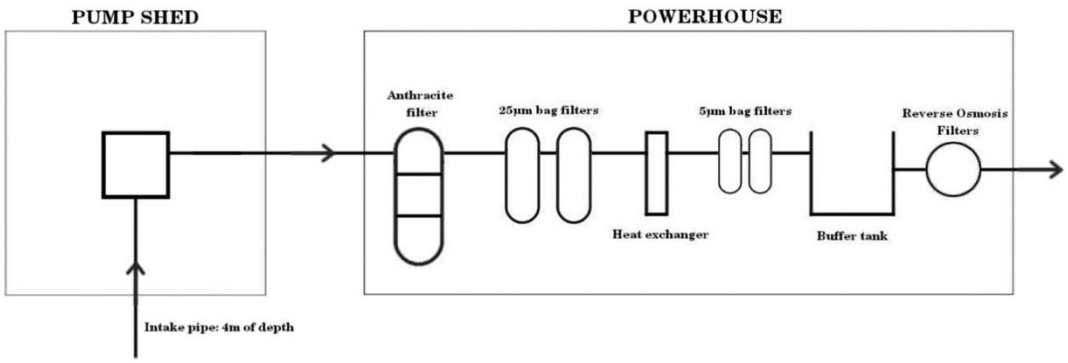


Figure 3: Simplified diagram of MZS desalination plant.

Differently from Sewell and Jury (2009), where 100 µm filters are “reusable” and regenerated after having been in use for the same amount of time, at MZS Station, the 25 µm bag and 5 µm cartridge filters are disposable, hence discarded after use. Their smaller mesh size, in fact, makes any potential

regeneration process unsuitable. Collected plankton samples for analyses are thus not obtained by washing the filters as in Sewell and Jury (2009), but only by the disruption of the filter structure (see below). In our case each filter is changed when the pressure inside the cartridge filter housing reached high levels, meaning that similar levels of plankton biomass and particulate matter are collected, regardless the amount of filtered seawater or time of activity, although this datum is always recorded in the logbook.

3.3.3. Sampling and laboratory procedures

Sampling was carried out in January and February of 2012 and 2013 enabling the collection of a total of eleven 5 µm cartridge filters, five in 2012 and six in 2013. The starting day for the two time ranges refers to the day in which the first filter was installed, differently from the sample name, which identifies the day it was sampled. For example, the filter "30_1_12" sampled the 30th of January of 2012 was installed 115 hours earlier, thus the 25th of January is the starting day for the time range investigated during 2012. The volume of water treated by the filters, based on data from the desalination plant electronic logbook, ranged from a minimum of 12.7 to a maximum of 64.8 m^3, with an average of approximately 23.41 m^3 per filter. The sea was in ice-free conditions from at least ten days before our sampling (Illuminati et al., 2017; Monti et al., 2017 for 2012; Schiaparelli personal communication for 2013). As soon as the pressure inside the cartridge filter housing reached high levels, the desalination plant technician informed one of the authors (SS) of the imminent replacement and let all the remaining seawater in the housing to flow "downstream" to the next desalination step. At this point filters were removed from the housing using lab gloves, placed in a sterile plastic bag and then stored at -20° C. These filters (Fig. 4a), measuring 50.8 cm of length and 6.4 cm of diameter, were kept at -20° C until summer 2018, when they were processed for the molecular analyses. Three replicates were obtained from each filter (one at the top, one in the middle and one at the end of the filter in order to cover all its length, see Fig. 4a), for a total of 33 replicates.

A metal, cylindrical, autoclave-sterilized cork borer of 26.25 mm in diameter was used to carve a circular cut on the surface of the cartridge filter. Different subsampling protocols were attempted on unused filters weeks before processing the filters used for this study, and tested by evaluating the amount and quality of the extracted DNA. During this optimization of the subsampling protocol, the deepest layers were found to yield a low amount of DNA. The most exterior layer of the filter (< 1 mm) was peeled off using a pair of heat-sterilized tweezers, in order to avoid any potential risk of post-sampling contamination, and discarded. Molecular analyses were thus performed on the immediately lower layer of the filter, and multiple cuts were performed for each replicate on different sides of the filter, enabling the extraction of the appropriate amount of sample weight required by most DNA extraction kits (*i.e.* at least the 0.25 g for the DNeasy PowerSoil Kit), also optimizing the amount of recoverable DNA.

Figure 4: (a) A frozen cartridge filter sampled on February 4th after having filtered ~22.5 hours. The three replicates were sampled from both extremities and the centre (blue arrows). (b) Layers of polypropylene extracted using a cork borer and a pair of heat-sterilized tweezers prior to the DNA extraction. Successively, the layers were cut in half and then in stripes of 1 mm of width. From Cecchetto et al. (2021).

3.3.4. Molecular analyses

Filter layers from each replicate were cut into small stripes (<1 mm) and then placed in the PowerBead Tubes provided by the DNeasy PowerSoil Kit (QIAGEN). DNA was extracted following the manufacturer's instructions, with the exception of an additional incubation step with the C1 solution

in a thermostatically controlled water bath (70° C for 10 minutes) and a final elution with 50 µl (instead of 100) of the C6 solution, in order to increase the DNA concentration. PCR amplification and sequencing of fragments of the 16S rRNA and 18S rRNA genes, for bacteria and eukaryotes respectively, were performed by IGA Technology (Udine, Italy, https://igatechnology.com/). The primers used for the V3 and V4 regions of 16S rRNA gene (approximately 450 bp) were chosen from Herlemann et al. (2011) and have the following sequences (Illumina adapters underlined): 341F - 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG 3' and 805R - 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC 3'. The primers used for the V4 and V5 regions of the 18S rRNA gene (approximately 550 bp) were selected from Hugerth et al. (2014) and have the following sequences (Illumina adapters underlined): 574*F – 5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCGGTAAYTCCAGCTCYV 3' and 1132R – 5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGTCAATTHCTTYAART 3'. The PCR mix was the same for both markers and consisted in 12.5 µl of 2x KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Woburn MA, USA), 5 µl of each primer and 2.5 µl of microbial DNA at a concentration of 5 ng/µl. The amplification conditions were: 95° C for 3 minutes, 25 cycles of 95° C for 30 seconds, 55° C for 30 seconds and 72° C for 30 seconds, followed by a final step at 72° C for 5 minutes.

A PCR clean-up step was performed using AMPure XP beads (Beckman Coulter) to purify from free primers and primer dimer species. This was followed by an indexing step using the Nextera XT Index (Illumina), to attach dual indices and Illumina sequencing adapters. The PCR program was the same of the amplicon PCR, except for the number of cycles set to 8 instead of 25. Another PCR clean-up step was performed prior to the quantification, normalization and sequencing using Illumina MiSeq v3 reagents on a 300 bp paired end reads MiSeq platform.

The PCR amplicons of the 16S rRNA region were sequenced on two different MiSeq runs to reach the minimum number of agreed sequences, which was 200,000 paired-end reads per replicate.

3.3.5. Bioinformatic analyses

Raw 18S rRNA sequences, after demultiplexing, were quality checked using *FastQC* and paired-end reads were merged using v*search* (Rognes et al., 2016), excluding merged products with more than 1 ambiguous base and more than 3 differences in the alignment. Primers were removed using c*utadapt* (Martin, 2011), allowing only one error in the alignment. *Mothur* (Schloss et al., 2009) was adopted to remove sequences with homopolymers greater than 8 bases, whereas *vsearch* was used to remove all sequences with a maximum expected error of 1, for length filtering (max 580 bp and min 490 bp) and for the dereplication. After the dereplication, the UNOISE2 algorithm (Edgar, 2016) implemented within *usearch* (Edgar, 2010), using the command "*unoise3*", was used to check for chimeras and remove singletons, generating the Zero-radius Operational Taxonomic Units (ZOTUs) fasta file. *Vsearch* was used again for the creation of a count table (command "*usearch_global*") using a global pairwise alignment with *id* equal to 1. The taxonomic assignment was conducted using the "Wang method" (naïve Bayesian classifier; Wang et al., 2007) implemented in *Mothur* and using version 4.12.0 of the PR2 database (Guillou et al., 2012).

Raw 16S rRNA sequences were processed with the same programs as for 18S rRNA, but with the following differences: the maximum differences allowed for merging were set to 10 (due to the longer alignment region for that primers), concatenation of the fastq files of the two different runs for each replicate, maximum expected error set to 0.5, length filtering set to 430 and 400 of maximum and minimum length respectively and the original (*i.e.* not modified) *mothur*-formatted version of the Silva database (release 132) for the taxonomic assignment (Quast et al., 2012).

The following bioinformatics analyses were all undertaken in R (version 3.6.3, R Core Team, 2020) and *Qiime2* (Bolyen et al., 2019). A variance stabilizing transformation, implemented in the R package *DESeq2* (Love et al., 2014) was applied to account for differences in the number of sequences, without prior merging of all the replicates. This stabilization was introduced as an alternative to the more common rarefaction method (McMurdie and Holmes, 2014). Negative values, which in the context of a variance stabilizing transformation indicate that in the original count table those values were more likely to be zero, or in any case negligible, were approximated to 0, as suggested by the *phyloseq* authors (McMurdie and Holmes, 2013) (https://www.bioconductor.org/packages/release/bioc/vignettes/phyloseq/inst/doc/phyloseq-FAQ.html#negative-numbers-in-my-transformed-data-table, last access on October 07 2020). This approximation allowed the calculation of Bray-Curtis distances for the ordination plot generated through a Non-metric Multidimensional Scaling (NMDS) using *phyloseq*. The Mantel test for evaluating a correlation between the distance matrices of 18S rRNA and 16S rRNA datasets was performed using *Qiime2*. Heatmaps were produced using the *phyloseq* R package, following the *phyloseq*-specific implementation of the NeatMap approach (Rajaram and Oono, 2010), adopting an ordination method instead of a hierarchical cluster analysis. Both heatmaps were calculated on Bray-Curtis distances and with a NMDS ordination. The heatmap for the 18S rRNA dataset was produced after reducing the count table to the 50th most abundant ZOTUs sorting samples by chronological order, from the 30th of January to the 5th of February of 2012 and from the 11th to the 25th of January of 2013. The heatmap for the 16S rRNA dataset was produced after reducing the count table to the 1000th most abundant ZOTUs, agglomerating them at the order level (fourth taxonomic level of the Silva Database) and sorting samples by chronological order. Taxa barplots were generated using *phyloseq* from the original, not transformed, count table after collapsing together all the replicates in the respective samples.

3.3.6. Environmental data: air temperature, wind and chlorophyll

AWS data on surface air temperature and wind direction and velocity were obtained from the "MeteoClimatological Observatory at MZS and Victoria Land" of PNRA (www.climantartide.it), for the AWS "Eneide" (-74.6959°, 164.0921°), located approximately 820 meters from the desalination plant pump shed. Data were processed in R using the packages *oce* (Kelley and Richards, 2020), *signal* (signal developers, 2014), *tsibble* (Wang et al., 2020), *dplyr* (Wickham et al., 2019b) and *cowplot* (Wilke, 2020).

Satellite data on chlorophyll (Ocean Biology Processing Group, 2018a) and POC (Ocean Biology Processing Group, 2018b) concentrations were obtained from NASA's OceanColor Web site using the level-3 browser to extract daily and monthly climatology data (from October to February of each year). Data were extracted choosing the "Standard" product at a 4 km resolution grid and for the area with the following latitudinal and longitudinal bounding box: -74.5°, -75°; 163.5°, 165°. The downloaded mapped files were converted from the format NetCDF to "csv" (comma separated values) using *GDAL* (Geospatial Data Abstraction Library, GDAL/OGR contributors, 2020) and processed in R using the *ggplot2* package (Wickham, 2016).

3.4. Results and Discussion

3.4.1. Particulate matter and filter replacement rate

Logbook data on filtering activity for filter cartridges (5 μm) and bags (25 μm) from 2002 to 2019 showed a consistent decrease in filtering activity hours from October to February, resulting in a higher rate of filter replacement towards the end of the summer (Fig. 5a and b). This observed higher rate of filter replacement since the end of the summer could be due to two different reasons: i) an increase of the desalination plant activity because of the intensification of the logistic activities in the research station, or ii) a progressive increase of the particulate matter present in the seawater. However, it is clear that the decreased filtering time in summer is not due to the logistic and scientific activities as the daily water requirement shows no particular trend (Fig. 5e) while, on the contrary, there is a clear increase of chlorophyll and POC from October to February (Fig. 5c and d). A more detailed overview on the temporal dynamics of filter replacement rate, with hourly and daily recordings of environmental variables throughout the 2011-2012 and 2012-2013 opening seasons, is provided in the supplementary material (Appendix Figs. 1 and 2).

This means that when the phytoplanktic bloom takes place, the increased amount of biomass in the seawater progressively and comparably determines an increase in the filter replacement rate, with a dramatic transition from weeks of activity of a single filter to peaks of multiple changes of filters per day. This situation takes place every year during the Antarctic summer, in conjunction with the sea ice retreat and the occurrence of phytoplanktic blooms triggered by sympagic communities (Mangoni et al., 2009; Saggiomo et al., 2017). The distribution of blooms is rather patchy, being influenced by the seasonal extension and shape of the marginal ice-zone. This determines a mosaic of different planktic communities in the water column, each one characterized by a different taxonomic composition (Nuccio et al., 2000). Other environmental drivers, such as winds, may introduce other sources of variability, further affecting community dynamics (Brandini, 1993; Fitch and Moore, 2007;

Moline and Prezelin, 1996). The effect of winds is especially important in Antarctica due to the existence of high-energy winds, *i.e.* katabatic winds, whose pulses can be considered extreme events.

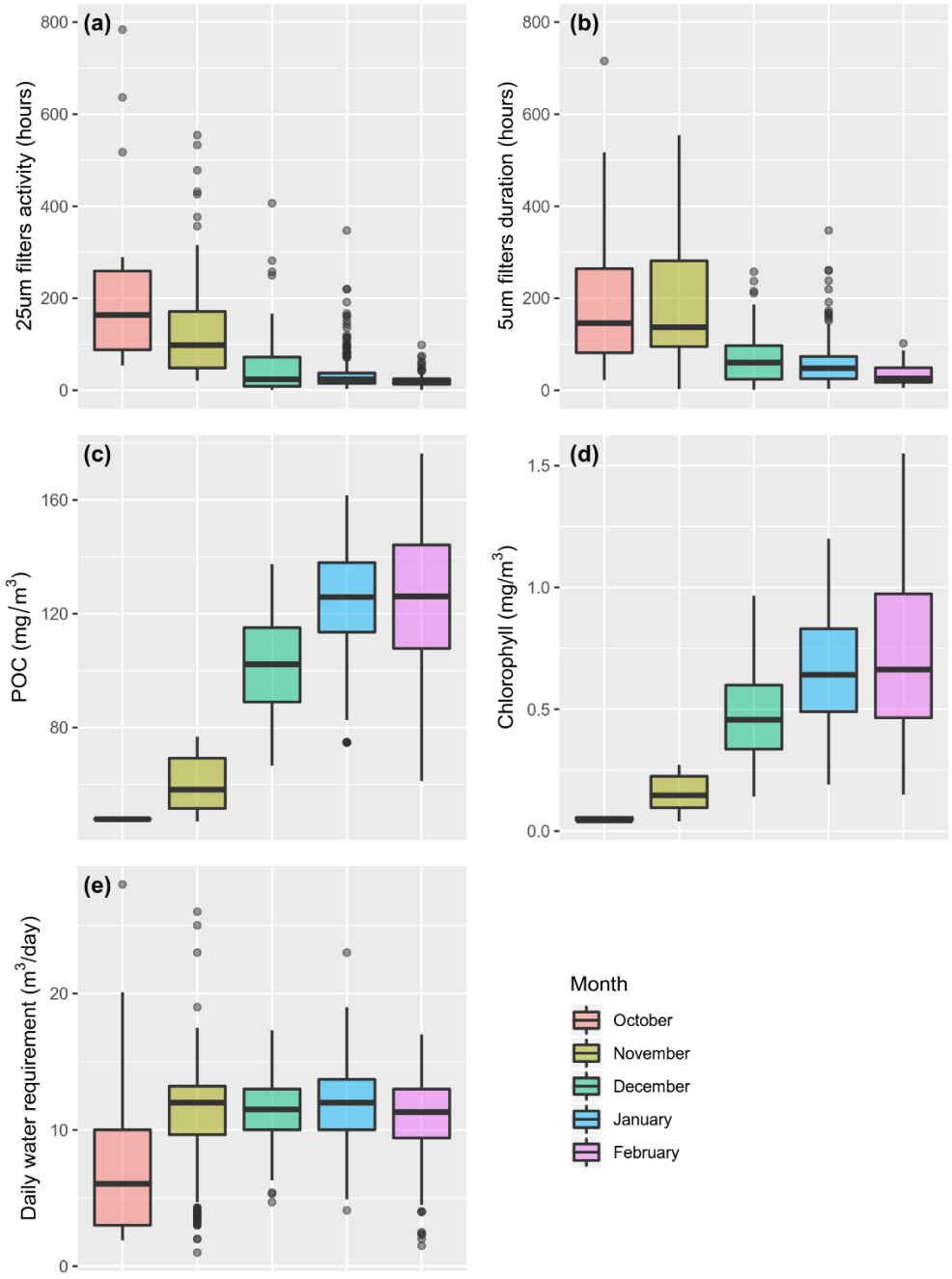


Figure 5: Boxplots of log-diary and satellite data from 2002 to 2019. Upper boxplots refer to (a) filter activity hours for bag filters (25 µm mesh size) and (b) cartridge filters (5 µm mesh size). Boxplots in the middle refer to (c) Particulate Organic Carbon (POC) and (d) Chlorophyll concentration measured in milligrams per cubic meter. Lower boxplot (e) refers to the total monthly cubic meters of water consumed by the research station. All data has been gathered based on month of registration and ordered from October to February. Satellite data for Chlorophyll and POC in October are less abundant than for the other months, as most of the area is usually covered in sea-ice during that period. Filters were assigned to month on the base of their installation time. From Cecchetto et al. (2021).

Thus, due to the high community patchiness and the presence of major environmental drivers, the availability of a higher sampling frequency is mandatory in order to unravel intra and inter-annual planktic dynamics, especially when it is known that rapid short-term variations have a high probability of occurrence as also shown by our data.

3.4.2. Community composition, diversity and dynamics of nanoplankton revealed by DNA metabarcoding

Bioinformatics analyses produced a total of 603 ZOTUs for 18S rRNA and 3,914 ZOTUs for 16S rRNA. Final abundance values add up to 1,219,853 and 1,726,680 sequences, corresponding to ~30% and ~38% of the total "raw" sequences for the 18S rRNA and 16S rRNA datasets, respectively.

The NMDS (Fig. 6) showed the ability of amplicon sequencing to differentiate nanoplanktic communities investigated during similar seasons of two consecutive years and to track short-term changes in community composition taking place in just a few days (Fig. 6). For both years the ordination showed a clear distinction between the first days and the following ones, meaning that the investigated time frame was characterized by a transition of the community composition from a particular state to another one. The same transition has been recorded both in the 16S rRNA and 18S rRNA datasets (Fig. 6), suggesting that the different environmental and biological conditions similarly influenced both communities, with a very neat and strict positive correlation between the two Bray-Curtis distance matrices (Pearson r=0.90387, p=0.001) (Fig. 7). Thus, any change in community composition might be tracked by DNA metabarcoding using alternatively 16S rRNA or 18S rRNA, which provide highly overlapping metrics.

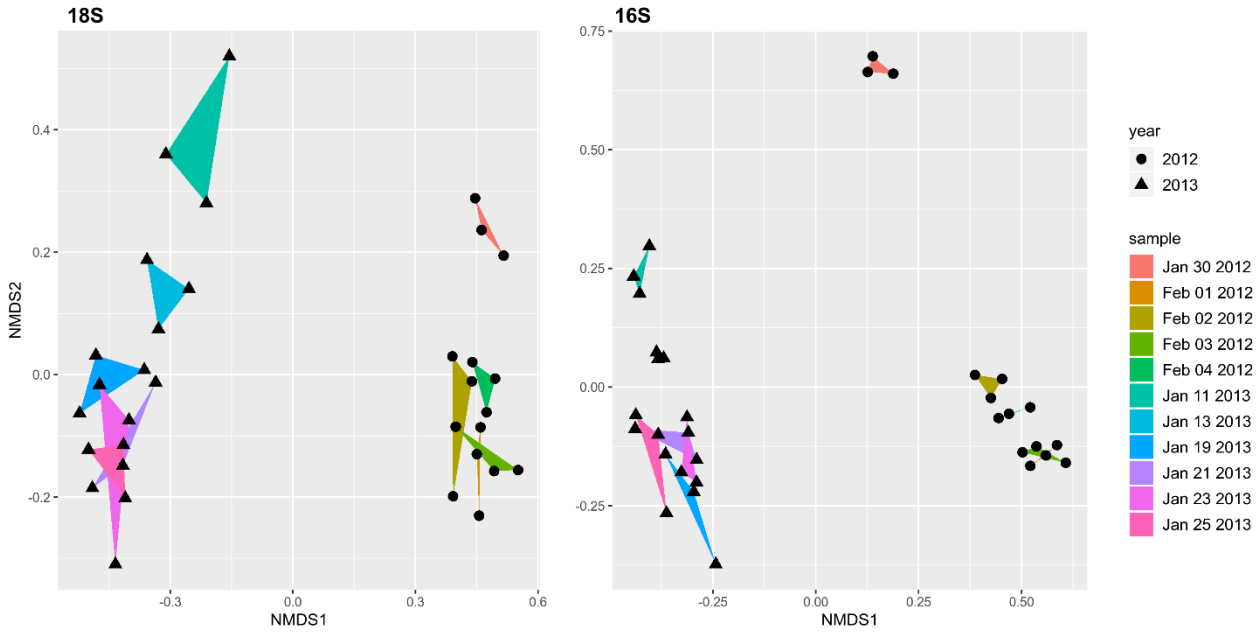


Figure 6: Non-metric Multidimensional Scaling of 16S and 18S based on Bray-Curtis distances. Colours refer to the replicates of the same filter, thus corresponding to the same day of sampling. Dates in the legend are ordered in temporal succession. Triangles refer to 2012 samples and circles to 2013 samples. From Cecchetto et al. (2021).

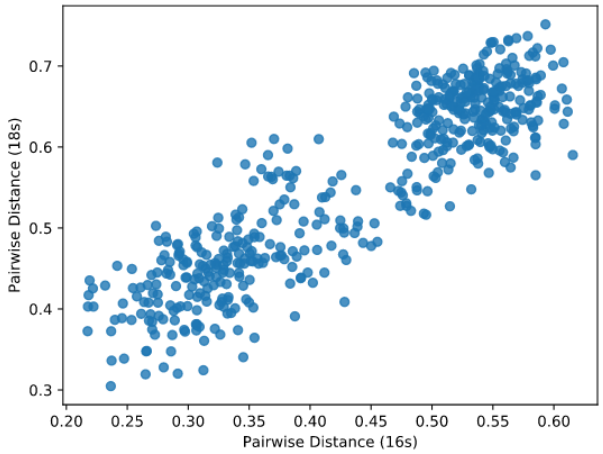


Figure 7: Scatterplot showing the correlation between the two different matrices of Bray-Curtis distances for 18S and 16S. From Cecchetto et al. (2021).

The nanoeukaryotic community here investigated showed a marked presence of different taxonomic groups of Dinophyceae in both years. The 2012 dataset was characterized by the presence of Gymnodiniales and a more relevant incidence of Metazoa (Arthropoda, Maxillopoda) and Suctoria (Ciliophora, Phyllopharyngea), while in January 2013 two unidentified groups of Dinophyceae resulted to be the most abundant taxa (Fig. 8a). The 16S rRNA dataset showed a community

resembling the typical composition of surface waters Antarctic copiotrophic prokaryotes, being dominated by the classes Bacteroidia, Alphaproteobacteria and Gammaproteobacteria, already evidenced in previous studies (Celussi et al., 2010; Giudice and Azzaro, 2019; Lo Giudice et al., 2012) (Fig. 8b). However, due to the mesh size of the cartridge filters, the bacterioplankton community here investigated should not be referred to free-living bacteria, but rather to particle-attached prokaryotes.

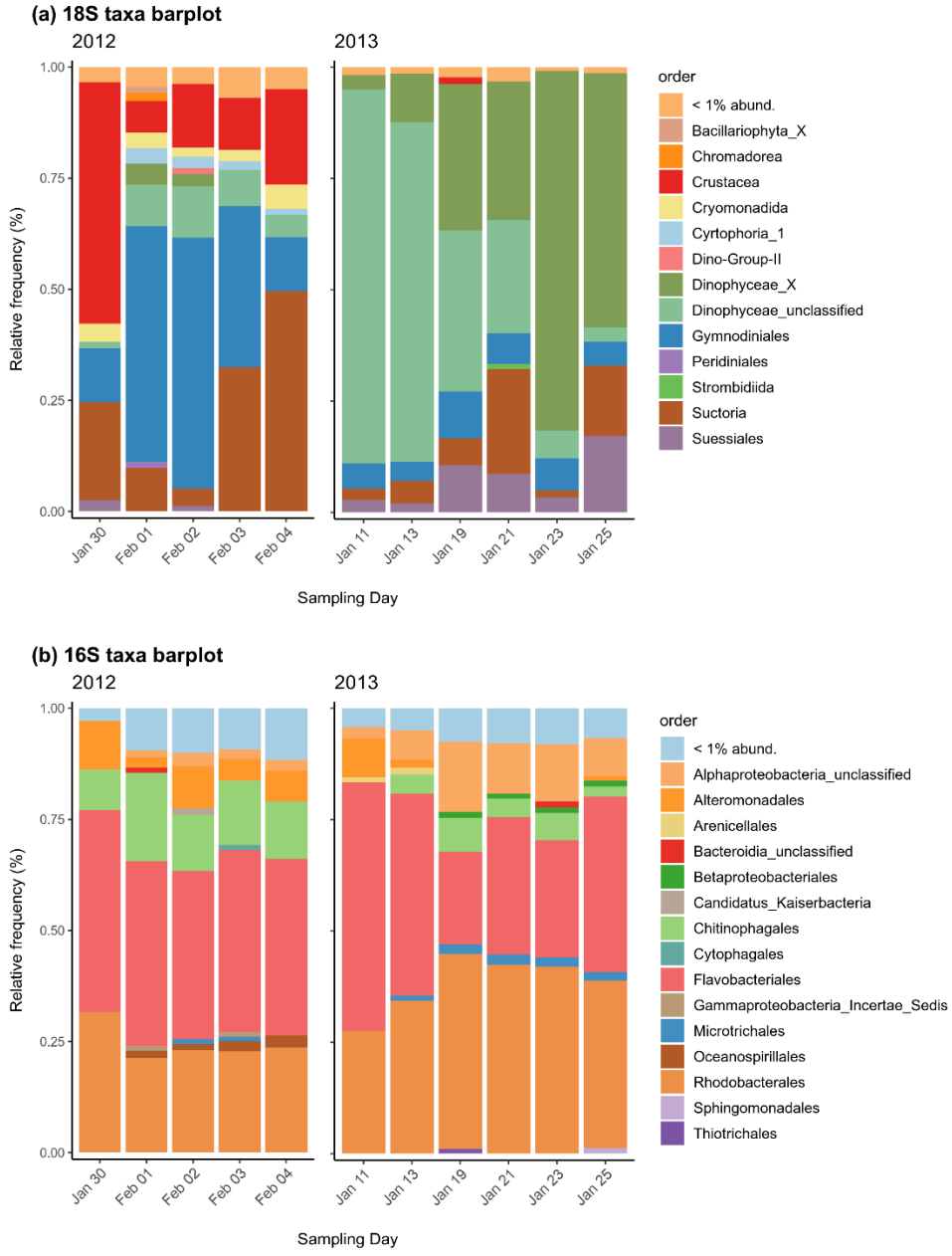


Figure 8: a) Taxa barplots for 18S of 2012 (left) and 2013 (right); b) Taxa barplots for 16S of 2012 (left) and 2013 (right). From Cecchetto et al. (2021).

The two different time ranges investigated also indicate different intra-annual dynamics: in 2012 there was a temporary (lasting only three days, from the 1th to the 3th of February) increase of the relative frequency of Gymnodiniales and other orders of Dinophyceae over Maxillopoda and Suctoria, while in 2013 there was a clear shift between two distinct groups of Dinoflagellates, represented in particular by the class Dinophyceae (Fig. 8a).

The abrupt change in the community composition detected in 2012 may be the result of a water column instability induced by katabatic wind pulses that, as shown by the AWS “Eneide” data (Fig. 9a and c), characterized two distinct periods of high wind intensity separated by an interval of 3 days with low-intensity winds (January 29th - February 1th). During this brief period of calm weather and water column stability there was an increase of different groups of Dinoflagellates, but this did not lead to a monospecific bloom, which was likely disrupted by the second katabatic event.

The high presence of Maxillopoda sequences registered during periods of high wind intensity should not be considered as indicative of the presence of crustacean adults on the cartridge filter itself, but as a possible result of spawning or molting events or even from disrupted body parts originating from individuals intercepted from upstream components of the desalination plant. The latter would be more likely, especially if we consider the equally high presence of Suctorian sequences, which are the most widespread symbiotic group in the phylum Ciliophora and can be found as facultative ectosymbionts on crustaceans (Lynn, 2008) or even on phytoplankton (Sazhin et al., 2007).

On the other hand, no katabatic event was recorded by the AWS during the 2013 time series, resulting in a relatively calm period with just sporadic peaks of wind intensity from different directions (Fig. 9b and d). These more stable conditions may have favoured a progressive shift between two different Dinoflagellate groups, without abrupt changes as those observed in the 2012 series.

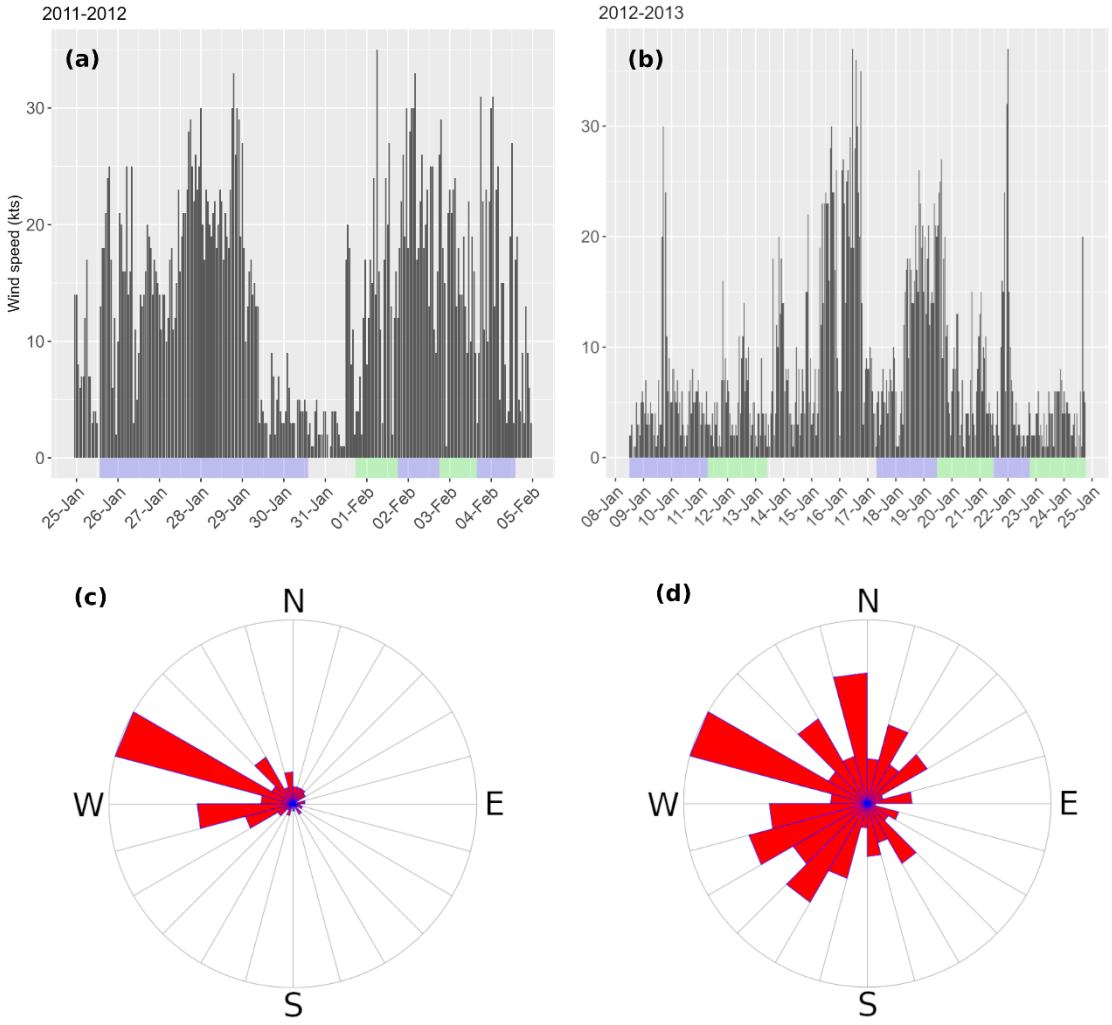


Figure 9: Barplots of hourly wind speed recordings (a,b) and wind roses (c,d) for the two different time ranges investigated during 2012 and 2013. (a) The wind intensity (knots) and c) wind rose at the bottom for the 2012 series, (b) and (d) for the 2013 series. Blue and green bands below the barplots indicate activity time ranges for the individual filters sampled for this study, whereas the grey areas represent the activity time of filters that couldn't be sampled for this study. From Cecchetto et al. (2021).

Regarding the bacterioplankton community of the 2013 series, the distinction in the community composition (Fig. 6) is likely to be determined by an increase in the relative abundance of Rhodobacterales (Alphaproteobacteria) and an unidentified group of Alphaproteobacteria, mostly to the detriment of Gammaproteobacteria and of Flavobacterales (Bacteroidia) (Fig 8b).

For the 2012 series, no evident temporal dynamics at higher taxonomic levels can be appreciated and the major difference inferred by the NMDS is likely to be the result of an abrupt increase in alpha diversity between the first day (January 30th) and the following ones (Fig. 10). The orders

Chitinophagales (Bacteroidia) and Alteromonadales (Alphaproteobacteria) were the only ones showing a slight increase and decrease (Fig. 8b) in percentage, respectively, probably reflecting the increase in phytoplanktic activity and algal-derived polymeric substrates (Wilkins et al., 2013) of both dinoflagellates and diatoms.

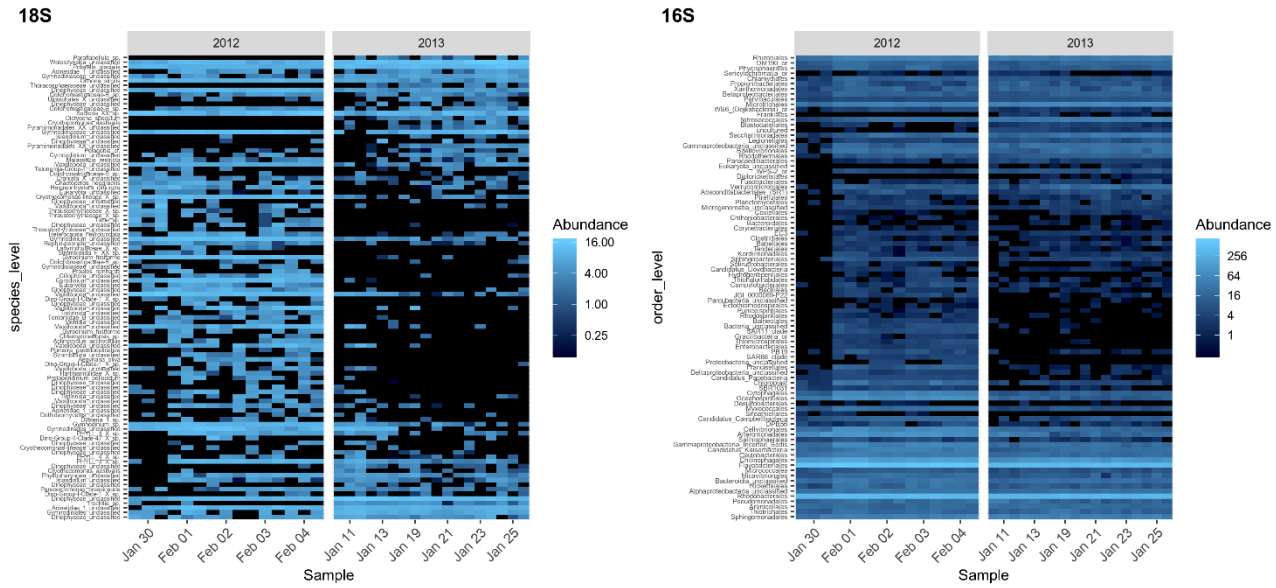


Figure 10: Heatmap of 18S 50 most abundant ZOTUs (left) and 16S 1,000 most abundant ZOTUs (right), the latter agglomerated by taxonomic order (fourth level from the highest of the Silva database taxonomy). Abundances values refer to those given after the variance stabilizing transformation. The x-axis is sorted in chronological order, from the oldest to the most recent filter, with the 2012 series on the left and the 2013 on the right, for both heatmaps. From Cecchetto et al. (2021).

The correlation between low wind activity and the development of phytoplanktic blooms has already been recognized, not only in Antarctic coastal planktic communities (*e.g.* Brandini, 1993; Moline and Prezelin, 1996), but also in Antarctic offshore areas (*e.g.* Kanta et al., 2017; Park et al., 2019; Sallée et al., 2015), as well as in non polar areas (*e.g.* Nieblas et al., 2009; Qu et al., 2020). High wind activity has a direct effect on the water column structure, being capable of mixing it and inducing upwelling phenomena, thus hampering bloom occurrences (Tripathy and Jena, 2019).

Our data showed a sudden temporal response of these communities after the reduction in wind intensity, which may have allowed a temporary condition of stability that, in turn, enabled the start of

a water column stratification process. This was reflected in the increase in dinoflagellate abundance, which, however, couldn't last more than three days due to the occurrence of a second katabatic event.

Unfortunately, most of the research based on HTS methodologies conducted in Terra Nova Bay (and especially near the Italian Research station "MZS") focused on prokaryotic communities only (Giudice and Azzaro, 2019) and not on the eukaryotic ones. Consequently, the absence of an in depth knowledge of coastal eukaryotic communities studied through metabarcoding hampers a critical comparison with our results. Nonetheless, the data obtained in this study showed a clear dominance of dinoflagellates in the nanoplanktic community, in accordance to what is known from previous study focusing on deeper water strata (Zoccarato et al., 2016) or on the sea ice (Torstensson et al., 2015), suggesting that these groups may play a very important and general role in Antarctic ecosystems (Liu et al., 2020). The absence of other protists, such as Radiolaria, Hacrobia and Excavata (which, apart from the latter, are nonetheless represented by some ZOTUs in the dataset), may be due to the difference in the size range investigated, wider in the aforementioned studies or simply to the intrinsic differences in the water masses examined or in the timing of sampling.

Several highly represented taxa in our results have never or only just rarely been documented in this area before. This is the case of some groups of eukaryotes such as: i) Cryomonadida (Cercozoa; Filosa-Thecofilosea), which graze on bacteria and may also parasitize phytoplankton (Zoccarato et al., 2016); ii) Cyrtophoria (Ciliophora; Phyllopharyngea), typically found in biofilms or as facultative or obligate symbionts on the body surfaces of invertebrates, such as crustaceans (Lynn, 2008); and iii) Suctoria (Ciliophora; Phyllopharyngea), this latter one representing the third most abundant taxon in the entire 18S rRNA dataset. The absence of diatoms in high number is a more surprising result, and will be discussed in the next section.

However, regarding the comparison of inter-annual dynamics, it has to be stressed that no conclusions can be drown despite the sampling activities occurred in the same season and with a similar timing. In fact, the investigated time ranges are too short and it is not possible to assess whether or not the two observed situations represent “typical” seasonal dynamics, just shifted in time.

3.4.3. Advantage of the method and possible implementations

Despite only few filters were available for this study, also limited only to the 5 µm mesh size fraction, the adopted molecular approach enabled a high-resolution analysis of intra-annual dynamics of Terra Nova Bay plankton. This was obtained through a cost-effective method (no funds were needed to set up the filtering system as it is part of the research base) and, especially, without the need of personnel at sea for continuous samplings, which is logistically unfeasible especially during extreme weather conditions that characterize katabatic wind events. Several new taxa were also recorded for the first time and future studies will enable clarifying if these are regular occurrences in the area.

Unfortunately, not enough filters have been studied so far in order to address the capacity of recovering rare taxa based on the different amount of filtered seawater, as most of the filters were in use for a similar amount of hours (see materials and methods). Sewell and Jury (2009) stated that the system is capable of recovering most of the rare taxa, but their methodology allowed a sampling frequency based on the quantity of seawater filtered, whereas in our study this approach would be logistically unfeasible. Only a couple of filters (the first of the two series) had significantly higher values of filtered seawater but, as mentioned earlier, they also were those with the lower numbers of taxa recorded.

Moreover, as the amount of particulate present in the input seawater does not necessarily correspond to higher biomass, uncertainties in the interpretation of actual bloom events may arise. This issue could be easily resolved by monitoring also other environmental and biological parameters (*e.g.*

turbidity and chlorophyll concentration) by establishing an *in situ* monitoring station located in the vicinity of the seawater intake pipe to obtain environmental data in continuum. The availability of these data will enable a more precise interpretation of the community changes disclosed by metabarcoding.

It has also to be considered that in metabarcoding studies the abundances of taxa are always difficult to be estimated in "absolute" terms for a variety of reasons (Taberlet et al., 2018), above all the well-known issues regarding primer amplification biases (Jovel et al., 2016; Piñol et al., 2019). The adoption of different methodologies, such as metagenomics (adopting shotgun sequencing techniques), which don't rely on amplification enrichment, would certainly reduce the impact of these issues (Bohmann et al., 2014). In this context, biodiversity monitoring using filters from desalination plants, and its usefulness in detecting short-term dynamics in coastal communities, would greatly benefit from the potentials of methodologies such as metatranscriptomics. In general, further and specific research would be required to validate the applicability of different methodologies, also according to the taxonomic group of interest, the project goals, and the availability of *in situ* lab facilities

Some eukaryotes recorded in this study are typically found growing on biofilms, such as Cryomonadida, which has already been documented in water treatment systems (Angell et al., 2020; Fried et al., 2000) and whose abundance could potentially result overestimated (Henthorne and Boysen, 2015). However, the dynamics of eukaryotic communities inside desalination plants are largely unknown and very few papers deal with this issue (*e.g.* Belila et al., 2017), the main focus of seawater pre-treatment studies having been bacterial biofilm eradication to prevent membrane clogging (*e.g.* Bar-Zeev et al., 2009). Nonetheless, due to the long period of inactivity of the Italian research station "MZS" during the winter, as well as the frequent replacement of different pre-treatment filters during most of the summer, the impact of potential biofilm growth should be

minimal. As stated before, in fact, the desalination plant is also closed at the end of each expedition by pumping air in all pipes and valves, completely removing the amount of liquid seawater at the end of each expedition. This cleaning practice, together with the high amount of seawater usually filtered daily through the entire desalination plant (which is fully replaced every 4 minutes) suggests as this contribution, although not quantifiable, should be really negligible. Thus, the data reported should really reflect what is present in the water column.

Surprisingly, diatoms, despite being usually reported as one of the main components of the phytoplanktic blooms (Mangoni et al., 2009; Pabi and Arrigo, 2006), were not abundant in our samples. Another survey, carried out during the austral summer 2011-2012, before and immediately after our sampling time frame at an offshore site, roughly 1 Km far from the desalination plant intake pipe, showed a dominance of diatoms, both in terms of cell abundances and biomass, while Dinoflagellates represented only a minor group (Illuminati et al., 2017). However, the method adopted by Truzzi et al. (2015) and Illuminati et al. (2017) involved a completely different protocol based on a quali-quantitative methodology, and not based on a selective filtration process. For this reason, the absence of abundant diatom sequences in the 5 µm “cartridge” filters may simply be due to the retention of most diatom species by the 25 µm “bag” filters located upstream. On the other hand, this apparent incongruence could also simply be due to the well-known patchy distribution of plankton communities in Terra Nova Bay, where areas dominated by diatom blooms are intermixed with others mainly dominated by dinoflagellates and other flagellates, also forming strong inshore-offshore gradients (Nuccio et al., 2000). Another reason could be related to the sub-sampling protocol and DNA extraction I have adopted. The chosen primers (Hugerth et al., 2014) should theoretically amplify 18S rRNA from Ochrophyta, as running an in silico PCR on the Silva SSU RefNR Database (Klindworth et al., 2013), allowing only two mismatches, reports 98% of coverage for that group. However, since the first layers of the filters were discarded and no aggressive steps, such as the

mechanical lysis of diatom cell wall (frustules), were adopted during DNA extraction (Vasselon et al., 2017), it is possible that the diatom component in the total DNA extract was potentially reduced.

A general aspect to consider for the proposed method would also be the storage conditions for the samples which, in this case, correspond to a storage at -20° C for some years. It is known that, despite being one of the most widespread techniques for storing samples used for molecular analyses, freezing at -20° C could be optimal for short periods of time, whereas, on the long-term, -80° C would be preferable (Straube and Juen, 2013). Other storage conditions were also proposed in literature, each one with different pros and cons that may condition the results of a study (*e.g.* Ransome et al., 2017) thus hampering comparisons between studies that adopt different storage protocols. In our case however, as samples were stored and processed under the same conditions, the comparison of observed dynamics are valid, potential biases being exactly the same for the two sets of samples.

An implementation of the method is the adoption of a subsampling procedure done immediately after filters' collection. This step greatly reduces the size of the samples (*i.e.* small cores instead whole filters have to be preserved) and also allows adopting different storage procedure (*e.g.* medium-based instead frozen). This simple step surely facilitates the storage and shipping of samples by greatly reducing their physical volume. Thanks to the increasing availability, portability and cost-efficiency of new molecular technologies (Gilbert, 2017; Johnson et al., 2017) all these analyses could also be ideally done in the field, thus completely eliminating storage-related potential issues or biases.

Thesis Conclusions

The consequences that environmental changes will have on biological communities in Antarctica are still difficult to be predicted, especially due to the synergistic, non-linear effects of multiple environmental factors and biological interactions, that further hamper our ability to draw meaningful conclusions on how these communities will response in the long-term. Despite the increasing collaboration and exchange of data and skills between researchers that have managed in the last decades at a global scale, there is still a great demand of long-term, globally coordinated programs, to effectively monitor, and link to specific environmental drivers those changes that are occurring in biological communities. The promising results obtained in biodiversity research performed through HTS technologies increased the attention of the scientific communities towards “DNA metabarcoding”, leading to the production of thousands of publications in the last decade that applied and tested these technologies to a variety of topics and with different purposes, without failing to mentions both the advantages and pitfalls.

The results presented in this study indicate how the last years of scientific research in Antarctica has sensibly increased the barcode completeness of metazoan reference libraries reducing the gap between species occurrences and sequence coverage, at least for an important area such as the Ross Sea MPA. However, the absence of a continuous evaluation and quantification of the barcode completeness hampers the capacity and will of the scientific community to focus their research on the least studied metazoan groups, thus limiting the impact of new initiatives undertaken to reduce this gap. By adopting a standardized approach which aggregates data from different, constantly updated biodiversity repositories, this gap can be quantified within a certain level of approximation, but nonetheless help in the identification of the most critical obstacles to sequencing the Antarctic biodiversity and providing up-to-date barcode reference libraries.

In this context, it becomes crucial to increase the quality standards of the available public data and provide a reliable link between biodiversity occurrences and corresponding museum specimen vouchers, with the potential to boost the fulfillment of this objective and strengthen the international collaboration between researchers and institutions of different nationalities and specializing in different taxonomic groups, geographic areas etc.

Different data repositories, such as the BOLD System, have already introduced very selective requirements upon the acceptance and upload of new sequence data, the most important being the uploading of multiple additional collateral data, with multimedia support that further reduce the distance between sequence and specimen vouchers (Troudet et al., 2018). Initiatives focusing on boosting the reduction of the gap have already been tested with interesting results (*e.g.* Hebert et al., 2013), and appear to become more and more feasible, in terms of both time and financial costs, thanks to the development of new molecular methodologies (*e.g.* Chang et al., 2020; Sire et al., 2019).

Long-term monitoring programs and experiments, aimed to properly identify changes in benthic assemblages and their drivers, are becoming more and more needed especially in those areas in which these changes are expected to have major and unpredictable consequences. However, such programs are particularly difficult to organize and conduct, especially in remote, harsh environments, which require a higher amount of resources and personnel. In order to provide a reliable instrument, capable of generating comparable results and provide meaningful information on the ongoing influence of environmental changes on Antarctic communities, such programs must also implement highly reproducible techniques and analyses providing quantifiable and reproducible results. The application of “DNA metabarcoding” to the study of both the planktic and benthic communities, as evidenced by this thesis, showed promising results in the application of such methodologies in a remote and isolated environment.

For benthic organisms, Autonomous Reef Monitoring Structures processed through metabarcoding represent a cost-effective and highly reproducible methodology. Their structural design and ease of use allows applying both traditional methodologies such as visual census of the spatial coverage of sessile fauna and new molecular techniques, thus helping to provide new quantifiable data, without renouncing to compare these results with a reference baseline.

The results obtained in this thesis showed how the community that colonized these structures is extremely similar to those that have been previously investigated in Antarctica on different artificial structures and in different geographical areas. The molecular analyses performed here allowed the comparison of the development of this community with data from studies performed outside of the Southern Ocean but adopting the same methodology, highlithing the great differences that arise in the colonization of the same substrata in different regions. By adopting these methodologies, the stresses and changes that biological communities will endure in the near future can be quantified worldwide. However, an extact quantification of these differences would require the adoption of the same sampling and bioinformatic protocols in multiple logistical contexts, a complex task to perform in harsh and logistically complicated environments like Antarctica. A strict collaboration between researchers and institutions would certainly overcome these difficulties, especially considering that monitoring programs at a continental and global scale are already occurring (Obst et al., 2020; Pearman et al., 2020).

“DNA metabarcoding” applied to the desalination plant filters, regardless of any technical peculiarity of a given desalination plant or mesh size considered, could represent a turning point in the always-increasing need of detailed and fine-scale data about the structure of phyto- and zooplankton inhabiting Antarctic coastal waters. Despite the need of further calibrations and the possible existence of issues that will require attention in the future, the availability of filters from a desalination plant offers unprecedented research opportunities at a more than achievable cost. Data shown here represent

a great leap in our knowledge of coastal plankton communities for the study area, highlighting previously unknown dynamics, such as the short-term and abrupt changes in coastal nanoeukaryotic communities' composition triggered by katabatic winds pulses, and finding groups of organisms never recorded before in an area (*i.e.* Terra Nova Bay) that was investigated by researchers for more than 30 years. This approach also overcomes most of the constraints linked to the logistic of sampling activities in a harsh environment and provides precise and fine-scale data that would simply not be achievable by using standard monitoring approaches based on the collection of water samples taken in the field, *e.g.* from the pack-ice or a boat.

By imagining a long term approach, where data of this type are collected each year at a given research station, it is out of doubts that the spatial and temporal dynamics of both the Antarctic coastal planktic and benthic communities will be revealed at unprecedented level of detail.
These data will be pivotal in our understanding of the complex dymanics occurring in Antarctica and will help in refining our predictions on possible future climatic scenarios.

Acknowledgements

First of all, I'd like to thank the two reviewers of this thesis, Diego Fontaneto and Cinzia Verde, for their insightful and very appreciated comments.

I'm also very grateful to the ENEA colleagues Francesco Pellegrino, Roberto Calvigioni and Fabio Baglioni for their collaboration and assistance during the 2018-2019 expedition in Antarctica.

I'd like to thank the Ocean Biology Processing Group (OBPG) at NASA's Goddard Space Flight Center for data availability at: https://oceancolor.gsfc.nasa.gov/. The PNRA "Meteo-Climatological Observatory" (http://www.climantartide.it) has kindly allowed the use of data regarding surface air temperature and wind direction and velocity.

I'm extremely grateful to my colleagues/friends, which always supported me since the beginning of my work at the Italian National Antarctic Museum. Thank you Claudio, Paola, Chiara, Alice, Marco and Guido. I've always appreciated the time spent with you all, I love you all and I wish you all the best.

The biggest thanks goes to Stefano Schiaparelli, my supervisor, who greatly supported me and without whom none of the work here reported would have been possible. I deeply admire his work, and the assistance (and patience) he gave me in all these years has been priceless. Thank you very much, Ste.

Most importantly, I'd like to thank my family and friends. Their support is the foundation to everything I am and what I've achieved in my life. I don't have words to express my love for all of you.

Everything that I love and have appreciated in my life was possible only thanks to my family, and for this reason I dedicate my thesis to my parents, my brother and my grandparents. I love you all.

Finally, I'd like to thank Ele. She constantly supported me, despite our personal ups and downs. Thank you for all our laughs.

References

A place for everything, 2008. Nature 453, 2. https://doi.org/10.1038/453002a

Adler, D., Murdoch, D., others, 2021. rgl: 3D Visualization Using OpenGL.

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'sullivan, C., Panchenko, A., Phan, L., Preuss, D., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E., Schneider, V., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E., Zbicz, K., 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. https://doi.org/10.1093/nar/gkv1290

Allcock, A.L., Strugnell, J.M., 2012. Southern Ocean diversity: New paradigms from molecular ecology. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2012.05.009

Amato, E., 1990. Environmental impact assessment at sea. Natl. Sci. Comm. Antarct. Oceanogr. Campaign 1987 88.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Bioinforma. Gr. Babraham Institute, Cambridge.

Angell, I.L., Bergaust, L., Hanssen, J.F., Aasen, E.M., Rudi, K., 2020. Ecological Processes Affecting Long-Term Eukaryote and Prokaryote Biofilm Persistence in Nitrogen Removal from Sewage. Genes (Basel). 11, 449. https://doi.org/10.3390/genes11040449

Antich, A., Palacín, C., Wangensteen, O.S., Turon, X., 2021. To denoise or to cluster? That is not the question. Optimizing pipelines for COI metabarcoding and metaphylogeography. bioRxiv.

Armbrecht, L.H., 2020. The potential of sedimentary ancient DNA to reconstruct past ocean ecosystems. Oceanography. https://doi.org/10.5670/oceanog.2020.211

Arranz, V., Pearman, W.S., Aguirre, J.D., Liggins, L., 2020. MARES, a replicable pipeline and curated reference database for marine eukaryote metabarcoding. Sci. Data. https://doi.org/10.1038/s41597-020-0549-9

Ashton, G. V., Morley, S.A., Barnes, D.K.A., Clark, M.S., Peck, L.S., 2017. Warming by 1°C Drives Species and Assemblage Level Responses in Antarctica's Marine Shallows. Curr. Biol. 27, 2698-2705.e3. https://doi.org/10.1016/j.cub.2017.07.048

Azevedo, J., Antunes, J.T., Machado, A.M., Vasconcelos, V., Leão, P.N., Froufe, E., 2020. Monitoring of biofouling communities in a Portuguese port using a combined morphological and metabarcoding approach. Sci. Rep. https://doi.org/10.1038/s41598-020-70307-4

Bacigalupi, L., Ramorino, M.C., 1994. Rapporto sulla Campagna Antartica Estate Australe 1993–94, Italian National Antarctic Program (PNRA).

Baird, D.J., Hajibabaei, M., 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next generation DNA sequencing. Mol. Ecol. 21, 2039–2044.

Balzano, S., Ellis, A. V, Le Lan, C., Leterme, S.C., 2015. Seasonal changes in phytoplankton on the north-eastern shelf of Kangaroo Island (South Australia) in 2012 and 2013. Oceanologia 57, 251–262. https://doi.org/https://doi.org/10.1016/j.oceano.2015.04.003

Bar-Zeev, E., Berman-Frank, I., Liberman, B., Rahav, E., Passow, U., Berman, T., 2009. Transparent exopolymer particles: Potential agents for organic fouling and biofilm formation in desalination and water treatment plants. Desalin. Water Treat. 3, 136–142. https://doi.org/https://doi.org/10.5004/dwt.2009.444

Barnes, D.K.A., 2013. Marine biology: New light on growth in the cold. Curr. Biol. 23, R609–R611. https://doi.org/10.1016/j.cub.2013.05.058

Barnes, D.K.A., Ashton, G. V., Morley, S.A., Peck, L.S., 2021. 1 °C warming increases spatial competition frequency and complexity in Antarctic marine macrofauna. Commun. Biol.

Barnes, D.K.A., Fenton, M., Cordingley, A., 2014. Climate-linked iceberg activity massively reduces spatial competition in Antarctic shallow waters. Curr. Biol. https://doi.org/10.1016/j.cub.2014.04.040

Barnes, D.K.A., Kuklinski, P., Jackson, J.A., Keel, G.W., Morley, S.A., Winston, J.E., 2011. Scott's collections help reveal accelerating marine life growth in Antarctica. Curr. Biol. https://doi.org/10.1016/j.cub.2011.01.033

Barnes, D. K., Fuentes, V., Clarke, A., Schloss, I. R., Wallace, M. I., 2006. Spatial and temporal variation in shallow seawater temperatures around Antarctica. Deep Sea Research Part II: Topical Studies in Oceanography, 53(8-10), 853-865. https://doi.org/10.1016/j.dsr2.2006.03.008

Bathmann, U., Fischer, G., Müller, P.J., Gerdes, D., 1991. Short-term variations in particulate matter sedimentation off Kapp Norvegia, Weddell Sea, Antarctica: relation to water mass advection, ice cover, plankton biomass and feeding activity. Polar Biol. 11, 185–195. https://doi.org/https://doi.org/10.1007/BF00240207

Belila, A., El-Chakhtoura, J., Saikaly, P.E., van Loosdrecht, M.C.M., Vrouwenvelder, J.S., 2017. Eukaryotic community diversity and spatial variation during drinking water production (by seawater desalination) and distribution in a full-scale network. Environ. Sci. Water Res. Technol. 3, 92–105. https://doi.org/https://doi.org/10.1039/C6EW00265J

Bindoff, N.L., Cheung, W.W.L., Kairo, J.G., Arstegui, J., Guinder, V.A., Hallberg, R., Hilmi, N., Jiao, N., Karim, M.S., Levin, L., 2019. Changing ocean, marine ecosystems, and dependent communities, in: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate.

Bogantes, V.E., Whelan, N. V., Webster, K., Mahon, A.R., Halanych, K.M., 2020. Unrecognized diversity of a scale worm, Polyeunoa laevis (Annelida: Polynoidae), that feeds on soft coral. Zool. Scr. https://doi.org/10.1111/zsc.12400

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y., De Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. 29, 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bowden, D.A., 2005. Quantitative characterization of shallow marine benthic assemblages at Ryder Bay, Adelaide Island, Antarctica. Mar. Biol. 146, 1235–1249. https://doi.org/10.1007/s00227-004-1526-0

Bowden, D.A., Clarke, A., Peck, L.S., Barnes, D.K.A., 2006. Antarctic sessile marine benthos: Colonisation and growth on artificial substrata over three years. Mar. Ecol. Prog. Ser. 316, 1–16. https://doi.org/10.3354/meps316001

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., Coissac, E., 2016. obitools: A unix inspired software package for DNA metabarcoding. Mol. Ecol. Resour. 16, 176–182.

Brandini, F.P., 1993. Phytoplankton biomass in an Antarctic coastal environment during stable water conditions-implications for the iron limitation theory. Mar. Ecol. Ser. 93, 267. https://doi.org/10.3354/meps093267

Brandt, A., De Broyer, C., De Mesel, I., Ellingsen, K.E., Gooday, A.J., Hilbig, B., Linse, K., Thomson, M.R.A., Tyler, P.A., 2007. The biodiversity of the deep Southern Ocean benthos. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2006.1952

Brannock, P.M., Learman, D.R., Mahon, A.R., Santos, S.R., Halanych, K.M., 2018. Meiobenthic community composition and biodiversity along a 5500 km transect of Western Antarctica: A metabarcoding analysis. Mar. Ecol. Prog. Ser. 603, 47–60. https://doi.org/10.3354/meps12717

Brasier, M.J., Grant, S.M., Trathan, P.N., Allcock, L., Ashford, O., Blagbrough, H., Brandt, A., Danis, B., Downey, R., Eléaume, M.P., Enderlein, P., Ghiglione, C., Hogg, O., Linse, K., Mackenzie, M., Moreau, C., Robinson, L.F., Rodriguez, E., Spiridonov, V., Tate, A., Taylor, M., Waller, C., Wiklund, H., Griffiths, H.J., 2018. Benthic biodiversity in the South Orkney Islands Southern Shelf Marine Protected Area. Biodiversity. https://doi.org/10.1080/14888386.2018.1468821

Brasier, M.J., Wiklund, H., Neal, L., Jeffreys, R., Linse, K., Ruhl, H., Glover, A.G., 2016. DNA barcoding uncovers cryptic diversity in 50% of deep-sea antarctic polychaetes. R. Soc. Open Sci. 3. https://doi.org/10.1098/rsos.160432

Buttigieg, P.L., Fadeev, E., Bienhold, C., Hehemann, L., Offre, P., Boetius, A., 2018. Marine microbes in 4D—using time series observation to assess the dynamics of the ocean microbiome and its links to ocean health. Curr. Opin. Microbiol. 43, 169–185. https://doi.org/10.1016/j.mib.2018.01.015

Cantone, G., Castelli, A., Gambi, M.C., 2000. Benthic Polychaetes off Terra Nova Bay and Ross Sea: Species Composition, Biogeography, and Ecological Role, in: Ross Sea Ecology. https://doi.org/10.1007/978-3-642-59607-0_40

Cantone, G., Di Pietro, N., 2001. Benthic littoral Polychaeta “Sedentaria” of Terra Nova Bay (Ross Sea, Antarctica). Antarct. Sci. https://doi.org/10.1017/S0954102001000025

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P.D., Vuillemin, A., Ariztegui, D., Arnaud, F., Belle, S., Bertilsson, S., Bigler, C., Bindler, R., Brown, A.G., Clarke, C.L., Crump, S.E., Debroas, D., Englund, G., Ficetola, G.F., Garner, R.E., Gauthier, J., Gregory-Eaves, I., Heinecke, L., Herzschuh, U., Ibrahim, A., Kisand, V., Kjær, K.H., Lammers, Y., Littlefair, J., Messager, E., Monchamp, M.-E., Olajos, F., Orsi, W., Pedersen, M.W., Rijal, D.P., Rydberg, J., Spanbauer, T., Stoof-Leichsenring, K.R., Taberlet, P., Talas, L., Thomas, C., Walsh, D.A., Wang, Y., Willerslev, E., van Woerkom, A., Zimmermann, H.H., Coolen, M.J.L., Epp, L.S., Domaizon, I., Alsos, I.G., Parducci, L., 2021. Lake Sedimentary DNA Research on Past Terrestrial and

Aquatic Biodiversity: Overview and Recommendations. Quaternary. https://doi.org/10.3390/quat4010006

Carr, C. M., Hardy, S. M., Brown, T. M., Macdonald, T. A., & Hebert, P. D., 2011. A tri-oceanic perspective: DNA barcoding reveals geographic structure and cryptic diversity in Canadian polychaetes. PLoS one, 6(7), e22232. https://doi.org/10.1371/journal.pone.0022232

Caruso, G., 2019. Microplastics as vectors of contaminants. Mar. Pollut. Bull. https://doi.org/10.1016/j.marpolbul.2019.07.052

Caruso, G., Azzaro, M., Dell'Acqua, O., Lo Giudice, A., Fazi, S., Caroppo, C., Azzaro, F., La Ferla, R., Maimone, G., Laganà, P., Marinelli, F., Berini, F., Marcone, G.L., Pichon, G., Chiantore, M., 2018. The ANT-biofilm project (PNRA): biological colonization of Antarctic coastal sites and biotechnological prospecting. Biol. Mar. Medit. 25, 267–268.

Carvalho, S., Aylagas, E., Villalobos, R., Kattan, Y., Berumen, M., Pearman, J.K., 2019. Beyond the visual: Using metabarcoding to characterize the hidden reef cryptobiome. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rspb.2018.2697

Cattaneo-Vietti, R., Chiantore, M., Gambi, M.C., Albertelli, G., Cormaci, M., Di Geronimo, I., 2000. Spatial and Vertical Distribution of Benthic Littoral Communities in Terra Nova Bay, in: Ross Sea Ecology. https://doi.org/10.1007/978-3-642-59607-0_36

CCAMLR, 2016. Conservation Measure 91-05.

Cecchetto, M., Di Cesare, A., Eckert, E., Fassio, G., Fontaneto, D., Moro, I., ... & Schiaparelli, S., 2021. Antarctic coastal nanoplankton dynamics revealed by metabarcoding of desalination plant filters: Detection of short-term events and implications for routine monitoring. Science of The Total Environment, 757, 143809. https://doi.org/10.1016/j.scitotenv.2020.143809

Cecchetto, M., Alvaro, M.C., Ghiglione, C., Guzzi, A., Mazzoli, C., Piazza, P., Schiaparelli, S., 2017. Distributional records of antarctic and sub-antarctic ophiuroidea from samples curated at the Italian national antarctic museum (MNA): Check-list update of the group in the terra nova bay

area (Ross Sea) and launch of the MNA 3D model ‘Virtual gallery.’ Zookeys 2017. https://doi.org/10.3897/zookeys.705.13712

Cecchetto, M., Lombardi, C., Canese, S., Cocito, S., Kuklinski, P., Mazzoli, C., Schiaparelli, S., 2019. The bryozoa collection of the italian national antarctic museum, with an updated checklist from Terra Nova Bay, Ross Sea. Zookeys 2019. https://doi.org/10.3897/zookeys.812.26964

Celussi, M., Bergamasco, A., Cataletto, B., Umani, S.F., Del Negro, P., 2010. Water masses’ bacterial community structure and microbial activities in the Ross Sea, Antarctica. Antarct. Sci. 22, 361. https://doi.org/10.1017/S0954102010000192

Celussi, M., Paoli, A., Crevatin, E., Bergamasco, A., Margiotta, F., Saggiomo, V., Umani, S.F., Del Negro, P., 2009. Short-term under-ice variability of prokaryotic plankton communities in coastal Antarctic waters (Cape Hallett, Ross Sea). Estuar. Coast. Shelf Sci. 81, 491–500. https://doi.org/10.1016/j.ecss.2008.12.014

Chain, F.J.J., Brown, E.A., MacIsaac, H.J., Cristescu, M.E., 2016. Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. Divers. Distrib. 22, 493–504. https://doi.org/10.1111/ddi.12427

Chamberlain, S., 2020a. worrms: World Register of Marine Species (WoRMS) Client.

Chamberlain, S., 2020b. bold: Interface to Bold Systems API.

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L., Ram, K., 2021. rgbif: Interface to the Global Biodiversity Information Facility API.

Chamberlain, S., Boettiger, C., 2017. R Python, and Ruby clients for GBIF species occurrence data. PeerJ Prepr.

Chang, J.J.M., Ip, Y.C.A., Bauman, A.G., Huang, D., 2020. MinION-in-ARMS: Nanopore Sequencing to Expedite Barcoding of Specimen-Rich Macrofaunal Samples From Autonomous Reef Monitoring Structures. Front. Mar. Sci. https://doi.org/10.3389/fmars.2020.00448

Chown, S. L., Clarke, A., Fraser, C. I., Cary, S. C., Moon, K. L., & McGeoch, M. A. (2015). The changing form of Antarctic biodiversity. Nature, 522(7557), 431-438. https://doi.org/10.1038/nature14505

Christiansen, H., Dettai, A., Heindler, F.M., Collins, M.A., Duhamel, G., Hautecoeur, M., Steinke, D., Volckaert, F.A.M., Van de Putte, A.P., 2018. Diversity of mesopelagic fishes in the Southern Ocean - A phylogeographic perspective using DNA barcoding. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2018.00120

Clarke, A., 2008. Antarctic marine benthic diversity: patterns and processes. Journal of Experimental Marine Biology and Ecology, 366(1-2), 48-55. https://doi.org/10.1016/j.jembe.2008.07.008

Clark, M.S., Villota Nieva, L., Hoffman, J.I., Davies, A.J., Trivedi, U.H., Turner, F., Ashton, G. V., Peck, L.S., 2019. Lack of long-term acclimation in Antarctic encrusting species suggests vulnerability to warming. Nat. Commun. https://doi.org/10.1038/s41467-019-11348-w

Compson, Z.G., McClenaghan, B., Singer, G.A.C., Fahner, N.A., Hajibabaei, M., 2020. Metabarcoding From Microbes to Mammals: Comprehensive Bioassessment on a Global Scale. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2020.581835

Contributors, G., 2020. GDAL/OGR Geospatial Data Abstraction Software Library.

Convey, P., Peck, L.S., 2019. Antarctic environmental change and biological responses. Sci. Adv. 5, eaaz0888. https://doi.org/10.1126/sciadv.aaz0888

Cornwell, W.K., Pearse, W.D., Dalrymple, R.L., Zanne, A.E., 2019. What we (don't) know about global plant diversity. Ecography (Cop.). https://doi.org/10.1111/ecog.04481

Cummings, V.J., Hewitt, J.E., Thrush, S.F., Marriott, P.M., Halliday, N.J., Norkko, A., 2018. Linking ross sea coastal benthic communities to environmental conditions: Documenting baselines in a spatially variable and changing world. Front. Mar. Sci. https://doi.org/10.3389/fmars.2018.00232

Czechowski, P., White, D., Clarke, L., McKay, A., Cooper, A., Stevens, M.I., 2016. Age-related environmental gradients influence invertebrate distribution in the prince Charles mountains, east Antarctica. R. Soc. Open Sci. https://doi.org/10.1098/rsos.160296

David, R., Uyarra, M.C., Carvalho, S., Anlauf, H., Borja, A., Cahill, A.E., Carugati, L., Danovaro, R., De Jode, A., Feral, J.P., Guillemain, D., Martire, M. Lo, D'Avray, L.T.D.V., Pearman, J.K., Chenuil, A., 2019. Lessons from photo analyses of Autonomous Reef Monitoring Structures as tools to detect (bio-)geographical, spatial, and environmental effects. Mar. Pollut. Bull. https://doi.org/10.1016/j.marpolbul.2019.02.066

Dayton, P., Jarrell, S., Kim, S., Thrush, S., Hammerstrom, K., Slattery, M., Parnell, E., 2016. Surprising episodic recruitment and growth of Antarctic sponges: Implications for ecological resilience. J. Exp. Mar. Bio. Ecol. https://doi.org/10.1016/j.jembe.2016.05.001

Dayton, P.K., 1989. Interdecadal variation in an antarctic sponge and its predators from oceanographic climate shifts. Science (80-.). https://doi.org/10.1126/science.245.4925.1484

Dayton, P.K., Jarrell, S.C., Kim, S., Ed Parnell, P., Thrush, S.F., Hammerstrom, K., Leichter, J.J., 2019. Benthic responses to an Antarctic regime shift: food particle size and recruitment biology. Ecol. Appl. 29, 1–20. https://doi.org/10.1002/eap.1823

Dayton, P.K., Kim, S., Jarrell, S.C., Oliver, J.S., Hammerstrom, K., Fisher, J.L., O'Connor, K., Barber, J.S., Robilliard, G., Barry, J., 2013. Recruitment, growth and mortality of an Antarctic hexactinellid sponge, Anoxycalyx joubini. PLoS One 8, e56939. https://doi.org/10.1371/journal.pone.0056939

De Broyer, C., Koubbi, P., Griffiths, H., Grant, S., 2014. Biogeographic Atlas of the Southern Ocean: Scientific Committee on Antarctic Research Cambridge.

Dettai, A., Adamowizc, S.J., Allcock, L., Arango, C.P., Barnes, D.K.A., Barratt, I., Chenuil, A., Couloux, A., Cruaud, C., David, B., Denis, F., Denys, G., Díaz, A., Eléaume, M., Féral, J.P., Froger, A., Gallut, C., Grant, R., Griffiths, H.J., Held, C., Hemery, L.G., Hosie, G., Kuklinski,

P., Lecointre, G., Linse, K., Lozouet, P., Mah, C., Monniot, F., Norman, M.D., O'Hara, T., Ozouf-Costaz, C., Piedallu, C., Pierrat, B., Poulin, E., Puillandre, N., Riddle, M., Samadi, S., Saucède, T., Schubart, C., Smith, P.J., Stevens, D.W., Steinke, D., Strugnell, J.M., Tarnowska, K., Wadley, V., Ameziane, N., 2011. DNA barcoding and molecular systematics of the benthic and demersal organisms of the CEAMARC survey. Polar Sci. https://doi.org/10.1016/j.polar.2011.02.002

DiTullio, G.R., Grebmeier, J.M., Arrigo, K.R., Lizotte, M.P., Robinson, D.H., Leventer, A., Barry, J.P., VanWoert, M.L., Dunbar, R.B., 2000. Rapid and early export of Phaeocystis antarctica blooms in the Ross Sea, Antarctica. Nature 404, 595–598. https://doi.org/10.1038/35007061

Douglass, L.L., Turner, J., Grantham, H.S., Kaiser, S., Constable, A., Nicoll, R., Raymond, B., Post, A., Brandt, A., Beaver, D., 2014. A hierarchical classification of benthic biodiversity and assessment of protected areas in the Southern Ocean. PLoS One. https://doi.org/10.1371/journal.pone.0100551

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv 81257. https://doi.org/10.1101/081257

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Edwards, M.E., 2020. The maturing relationship between Quaternary paleoecology and ancient sedimentary DNA. Quat. Res. (United States). https://doi.org/10.1017/qua.2020.52

Fabri-Ruiz, S., Danis, B., Navarro, N., Koubbi, P., Laffont, R., Saucède, T., 2020. Benthic ecoregionalization based on echinoid fauna of the Southern Ocean supports current proposals of Antarctic Marine Protected Areas under IPCC scenarios of climate change. Glob. Chang. Biol. https://doi.org/10.1111/gcb.14988

Faranda, F.M., Guglielmo, L., Ianora, A., 2000. The Italian Oceanographic Cruises in the Ross Sea (1987–95): Strategy, General Considerations and Description of the Sampling Sites, in: Ross Sea Ecology. https://doi.org/10.1007/978-3-642-59607-0_1

Ficetola, G.F., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., Taberlet, P., Pompanon, F., 2010. An in silico approach for the evaluation of DNA barcodes. BMC Genomics 11, 1–10.

Ficetola, G.F., Miaud, C., Pompanon, F., Taberlet, P., 2008. Species detection using environmental DNA from water samples. Biol. Lett. https://doi.org/10.1098/rsbl.2008.0118

Ficetola, G.F., Poulenard, J., Sabatier, P., Messager, E., Gielly, L., Leloup, A., Etienne, D., Bakke, J., Malet, E., Fanget, B., Støren, E., Reyss, J.L., Taberlet, P., Arnaud, F., 2018. DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Sci. Adv. https://doi.org/10.1126/sciadv.aar4292

Fillinger, L., Janussen, D., Lundälv, T., Richter, C., 2013. Rapid glass sponge expansion after climate-induced antarctic ice shelf collapse. Curr. Biol. 23, 1330–1334. https://doi.org/10.1016/j.cub.2013.05.051

Fitch, D.T., Moore, J.K., 2007. Wind speed influence on phytoplankton bloom dynamics in the Southern Ocean Marginal Ice Zone. J. Geophys. Res. Ocean. 112. https://doi.org/10.1029/2006JC004061

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. https://doi.org/10.1071/ZO9660275

Fontes, J.T., Vieira, P.E., Ekrem, T., Soares, P., Costa, F.O., 2021. BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. Mol. Ecol. Resour. https://doi.org/10.1111/1755-0998.13262

Foottit, R. G., Maw, H. E. L., Havill, N. P., Ahern, R. G., & Montgomery, M. E., 2009. DNA barcodes to identify species and explore diversity in the Adelgidae (Insecta: Hemiptera: Aphidoidea). Molecular Ecology Resources, 9, 188-195. https://doi.org/10.1111/j.1755-0998.2009.02644.x

Fried, J., Mayr, G., Berger, H., Traunspurger, W., Psenner, R., Lemmer, H., 2000. Monitoring protozoa and metazoa biofilm communities for assessing wastewater quality impact and reactor up-scaling effects. Water Sci. Technol. 41, 309–316. https://doi.org/10.2166/wst.2000.0460

Gao, C.-H., 2021. ggVennDiagram: A “ggplot2” Implement of Venn Diagram.

Garlaschè, Giuseppe, Karimullah, K., Iakovenko, N., Velasco-Castrillón, A., Janko, K., Guidetti, R., Rebecchi, L., Cecchetto, M., Schiaparelli, S., Jersabek, C.D., de Smet, W.H., Fontaneto, D., 2019. A data set on the distribution of Rotifera in Antarctica. Biogeographia. https://doi.org/10.21426/B635044786

Garlaschè, G., Karimullah, K., Iakovenko, N., Velasco-Castrillón, A., Janko, K., Guidetti, R., Rebecchi, L., Cecchetto, M., Schiaparelli, S., Jersabek, C.D., de Smet, W.H., Fontaneto, D., 2019. A data set on the distribution of Rotifera in Antarctica. Biogeographia 35. https://doi.org/10.21426/B635044786

Gast, R.J., Moran, D.M., Beaudoin, D.J., Blythe, J.N., Dennett, M.R., Caron, D.A., 2006. ABUNDANCE OF A NOVEL DINOFLAGELLATE PHYLOTYPE IN THE ROSS SEA, ANTARCTICA 1. J. Phycol. 42, 233–242. https://doi.org/10.1111/j.1529-8817.2006.00183.x

GBIF: The Global Biodiversity Information Facility, 2021. What is GBIF?

GBIF Occurrence Download https://doi.org/10.15468/dl.aq96re [WWW Document], n.d. https://doi.org/10.15468/dl.aq96re

GBIF Secretariat, 2021. GBIF Science Review 2020.

Ghiglione, C., Alvaro, M.C., Cecchetto, M., Canese, S., Downey, R., Guzzi, A., Mazzoli, C., Piazza, P., Rapp, H.T., Sarà, A., Schiaparelli, S., 2018. Porifera collection of the italian national antarctic museum (MNA), with an updated checklist from Terra Nova Bay (Ross Sea). Zookeys. https://doi.org/10.3897/zookeys.758.23485

Gilbert, M.T.P., 2017. Documenting DNA in the dust. Mol. Ecol. 26, 969–971. https://doi.org/10.1111/mec.13944

Gili, J. M., Coma, R., Orejas, C., López-González, P. J., & Zabala, M., 2001. Are Antarctic suspension-feeding communities different from those elsewhere in the world?. Polar Biology, 24(7), 473-485. https://doi.org/0.1007/s003000100257

Giudice, A. Lo, Azzaro, M., 2019. Diversity and ecological roles of prokaryotes in the changing Antarctic marine environment, in: The Ecological Role of Micro-Organisms in the Antarctic Environment. Springer, pp. 109–131.

Grant, R.A., Griffiths, H.J., Steinke, D., Wadley, V., Linse, K., 2011. Antarctic DNA barcoding; a drop in the ocean? Polar Biol. https://doi.org/10.1007/s00300-010-0932-7

Grant, R.A., Linse, K., 2009. Barcoding Antarctic biodiversity: Current status and the caml initiative, a case study of Marine invertebrates. Polar Biol. https://doi.org/10.1007/s00300-009-0662-x

Gratton P, Marta S, Bocksberger G, Winter M, Trucchi E, K hl H, 2017. A world of sequences: can we use georeferenced nucleotide databases for a robust automated phylogeography? J. Biogeogr.

Guillaumot, C., Martin, A., Eléaume, M., Saucède, T., 2018. Methods for improving species distribution models in data-poor areas: Example of sub-Antarctic benthic species on the Kerguelen Plateau. Mar. Ecol. Prog. Ser. https://doi.org/10.3354/meps12538

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., 2012. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41, D597–D604. https://doi.org/10.1093/nar/gks1160

Gutt, J., Isla, E., Xavier, J.C., Adams, B.J., Ahn, I.Y., Cheng, C.H.C., Colesie, C., Cummings, V.J., di Prisco, G., Griffiths, H., Hawes, I., Hogg, I., McIntyre, T., Meiners, K.M., Pearce, D.A., Peck, L., Piepenburg, D., Reisinger, R.R., Saba, G.K., Schloss, I.R., Signori, C.N., Smith, C.R., Vacchi, M., Verde, C., Wall, D.H., 2020. Antarctic ecosystems in transition – life between stresses and opportunities. Biol. Rev. https://doi.org/10.1111/brv.12679

Gwiazdowski, R.A., Foottit, R.G., Maw, H.E.L., Hebert, P.D.N., 2015. The Hemiptera (Insecta) of Canada: constructing a reference library of DNA barcodes. PLoS One 10, e0125635.

Hans, J., Jöst, C., Zauke, G.P., 2011. Significance and interspecific variability of accumulated trace metal concentrations in Antarctic benthic polychaetes. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2011.04.014

Hebert, P.D.N., Ratnasingham, S., DeWaard, J.R., 2003. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. B Biol. Sci. https://doi.org/10.1098/rsbl.2003.0025

Heimeier, D., Lavery, S., Sewell, M.A., 2010a. Using DNA barcoding and phylogenetics to identify Antarctic invertebrate larvae: Lessons from a large scale study. Mar. Genomics 3, 165–177. https://doi.org/10.1016/j.margen.2010.09.004

Heimeier, D., Lavery, S., Sewell, M.A., 2010b. Molecular species identification of Astrotoma agassizii from planktonic embryos: further evidence for a cryptic species complex. J. Hered. 101, 775–779. https://doi.org/10.1093/jhered/esq074

Hemery, L.G., Eléaume, M., Roussel, V., Améziane, N., Gallut, C., Steinke, D., Cruaud, C., Couloux, A., Wilson, N.G., 2012. Comprehensive sampling reveals circumpolarity and sympatry in seven mitochondrial lineages of the Southern Ocean crinoid species Promachocrinus kerguelensis (Echinodermata). Mol. Ecol. https://doi.org/10.1111/j.1365-294X.2012.05512.x

Henthorne, L., Boysen, B., 2015. State-of-the-art of reverse osmosis desalination pretreatment. Desalination 356, 129–139. https://doi.org/10.1016/j.desal.2014.10.039

Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., Andersson, A.F., 2011. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J. 5, 1571–1579. https://doi.org/10.1038/ismej.2011.41

Hernández-Triana, L. M., Prosser, S. W., Rodríguez-Perez, M. A., Chaverri, L. G., Hebert, P. D. N., & Ryan Gregory, T., 2014. Recovery of DNA barcodes from blackfly museum specimens (D

iptera: S imuliidae) using primer sets that target a variety of sequence lengths. Molecular Ecology Resources, 14(3), 508-518. https://doi.org/10.1111/1755-0998.12208

Holland, O., Shaw, J., Stark, J.S., Wilson, K.A., 2021. Hull fouling marine invasive species pose a very low, but plausible, risk of introduction to East Antarctica in climate change scenarios. Divers. Distrib.

Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C.B., Brandão, S.N., Gofas, S., Hooper, J.N.A., Hernandez, F., Holovachov, O., Mees, J., Molodtsova, T.N., Paulay, G., Decock, W., Dekeyzer, S., Lanssens, T., Vandepitte, L., Vanhoorne, B., Adlard, R., Adriaens, P., Agatha, S., Ahn, K.J., Akkari, N., Alvarez, B., Anderson, G., Angel, M. V, Antic, D., Arango, C., Artois, T., Atkinson, S., Bank, R., Barber, A., Barbosa, J.P., Bartsch, I., Bellan-Santini, D., Bernot, J., Berta, A., Bezerra, T.N., Bieler, R., Blanco, S., Blasco-Costa, I., Blazewicz, M., Bock, P., Böttger-Schnack, R., Bouchet, P., Boury-Esnault, N., Boxshall, G., Bray, R., Bruce, N.L., Cairns, S., Carballo, J.L., Cárdenas, P., Carstens, E., Chan, B.K., Chan, T.Y., Cheng, L., Churchill, M., Coleman, C.O., Collins, A.G., Collins, G.E., Corbari, L., Cordeiro, R., Cornils, A., Coste, M., Costello, M.J., Crandall, K.A., Cremonte, F., Cribb, T., Cutmore, S., Dahdouh-Guebas, F., Daly, M., Daneliya, M., Dauvin, J.C., Davie, P., De Broyer, C., De Grave, S., de Mazancourt, V., de Voogd, N.J., Decker, P., Decraemer, W., Defaye, D., D'Hondt, J.L., Dippenaar, S., Dohrmann, M., Dolan, J., Domning, D., Downey, R., Ector, L., Eisendle-Flöckner, U., Eitel, M., Encarnação, S.C. d., Enghoff, H., Epler, J., Ewers-Saucedo, C., Faber, M., Feist, S., Figueroa, D., Finn, J., Fišer, C., Fordyce, E., Foster, W., Frank, J.H., Fransen, C., Furuya, H., Galea, H., Garcia-Alvarez, O., Garic, R., Garnett, S., Gasca, R., Gaviria-Melo, S., Gerken, S., Gibson, D., Gibson, R., Gil, J., Gittenberger, A., Glasby, C., Glover, A., Gómez-Noguera, S.E., González-Solís, D., Gordon, D., Grabowski, M., Gravili, C., Guerra-García, J.M., Guidetti, R., Guiry, M.D., Hadfield, K.A., Hajdu, E., Hallermann, J., Hayward, B.W., Hendrycks, E., Herbert, D., Herrera Bachiller, A., Ho, J. s., Hodda, M., Høeg, J., Hoeksema, B., Houart, R., Hughes, L., Hyžný, M.,

Iniesta, L.F.M., Iseto, T., Ivanenko, S., Iwataki, M., Janssen, R., Jarms, G., Jaume, D., Jazdzewski, K., Jersabek, C.D., Jóźwiak, P., Kabat, A., Kantor, Y., Karanovic, I., Karthick, B., Kim, Y.H., King, R., Kirk, P.M., Klautau, M., Kociolek, J.P., Köhler, F., Kolb, J., Kotov, A., Kremenetskaia, A., Kristensen, R.M., Kulikovskiy, M., Kullander, S., Lambert, G., Lazarus, D., Le Coze, F., LeCroy, S., Leduc, D., Lefkowitz, E.J., Lemaitre, R., Liu, Y., Lörz, A.N., Lowry, J., Ludwig, T., Lundholm, N., Macpherson, E., Madin, L., Mah, C., Mamo, B., Mamos, T., Manconi, R., Mapstone, G., Marek, P.E., Marshall, B., Marshall, D.J., Martin, P., Mast, R., McFadden, C., McInnes, S.J., Meidla, T., Meland, K., Merrin, K.L., Messing, C., Miljutin, D., Mills, C., Moestrup, Ø., Mokievsky, V., Monniot, F., Mooi, R., Morandini, A.C., Moreira da Rocha, R., Moretzsohn, F., Morrow, C., Mortelmans, J., Mortimer, J., Musco, L., Neubauer, T.A., Neubert, E., Neuhaus, B., Ng, P., Nguyen, A.D., Nguyen Thi My, Y., Nielsen, C., Nishikawa, T., Norenburg, J., O'Hara, T., Opresko, D., Osawa, M., Osigus, H.J., Ota, Y., Páll-Gergely, B., Patterson, D., Paxton, H., Peña-Santiago, R., Perrier, V., Perrin, W., Petrescu, I., Picton, B., Pilger, J.F., Pisera, A.B., Polhemus, D., Poore, G.C., Potapova, M., Pugh, P., Read, G., Reich, M., Reimer, J.D., Reip, H., Reuscher, M., Reynolds, J.W., Richling, I., Rimet, F., Ríos, P., Rius, M., Rogers, D.C., Rosenberg, G., Rützler, K., Sabbe, K., Saiz-Salinas, J., Sala, S., Santagata, S., Santos, S., Sar, E., Satoh, A., Saucède, T., Schatz, H., Schierwater, B., Schmidt-Rhaesa, A., Schneider, S., Schönberg, C., Schuchert, P., Senna, A.R., Serejo, C., Shaik, S., Shamsi, S., Sharma, J., Shear, W.A., Shenkar, N., Shinn, A., Short, M., Sicinski, J., Sierwald, P., Simmons, E., Sinniger, F., Sivell, D., Sket, B., Smit, H., Smit, N., Smol, N., Souza-Filho, J.F., Spelda, J., Sterrer, W., Stienen, E., Stoev, P., Stöhr, S., Strand, M., Suárez-Morales, E., Summers, M., Suppan, L., Suttle, C., Swalla, B.J., Taiti, S., Tanaka, M., Tandberg, A.H., Tang, D., Tasker, M., Taylor, J., Taylor, J., Tchesunov, A., ten Hove, H., ter Poorten, J.J., Thomas, J.D., Thuesen, E. V, Thurston, M., Thuy, B., Timi, J.T., Timm, T., Todaro, A., Turon, X., Tyler, S., Uetz, P., Uribe-Palomino, J., Utevsky, S., Vacelet, J., Vachard, D., Vader, W., Väinölä, R., Valls Domedel, G.,

Van de Vijver, B., van der Meij, S.E., van Haaren, T., van Soest, R.W., Vanreusel, A., Venekey, V., Vinarski, M., Vonk, R., Vos, C., Walker-Smith, G., Walter, T.C., Watling, L., Wayland, M., Wesener, T., Wetzel, C.E., Whipps, C., White, K., Wieneke, U., Williams, D.M., Williams, G., Wilson, R., Witkowski, A., Witkowski, J., Wyatt, N., Wylezich, C., Xu, K., Zanol, J., Zeidler, W., Zhao, Z., 2021. World Register of Marine Species (WoRMS). https://doi.org/doi:10.14284/170

Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.12613

Huang, D., Meier, R., Todd, P.A., Chou, L.M., 2008. Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J. Mol. Evol. https://doi.org/10.1007/s00239-008-9069-5

Hugerth, L.W., Muller, E.E.L., Hu, Y.O.O., Lebrun, L.A.M., Roume, H., Lundin, D., Wilmes, P., Andersson, A.F., 2014. Systematic design of 18S rRNA gene primers for determining eukaryotic diversity in microbial consortia. PLoS One 9, e95567. https://doi.org/10.1371/journal.pone.0095567

Illuminati, S., Annibaldi, A., Romagnoli, T., Libani, G., Antonucci, M., Scarponi, G., Totti, C., Truzzi, C., 2017. Distribution of Cd, Pb and Cu between dissolved fraction, inorganic particulate and phytoplankton in seawater of Terra Nova Bay (Ross Sea, Antarctica) during austral summer 2011–12. Chemosphere 185, 1122–1135. https://doi.org/10.1016/j.chemosphere.2017.07.087

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., & Hebert, P. D., 2007. Universal primer cocktails for fish DNA barcoding. Molecular Ecology Notes, 7(4), 544-548. https://doi.org/10.1111/j.1471-8286.2007.01748.x

Jenkins, T.L., Stevens, J.R., 2018. Assessing connectivity between MPAs: Selecting taxa and translating genetic data to inform policy. Mar. Policy. https://doi.org/10.1016/j.marpol.2018.04.022

Johnson, S.S., Zaikova, E., Goerlitz, D.S., Bai, Y., Tighe, S.W., 2017. Real-time DNA sequencing in the Antarctic dry valleys using the Oxford Nanopore sequencer. J. Biomol. Tech. JBT 28, 2. https://doi.org/10.7171/jbt.17-2801-009

Jones, S., King, D., Robinson, S., Bergstrom, D., Waterman, M., 2020. Current state of Antarctic biodiversity monitoring. Conv. Abstr. Scar online 2020. https://www.scar2020.org/

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., 2016. Characterization of the gut microbiome using 16S or shotgun metagenomics. Front. Microbiol. 7, 459. https://doi.org/10.3389/fmicb.2016.00459

Kang, Y.H., Kim, S., Choi, S.K., Moon, K., Choi, H.G., Ko, Y.W., Hawes, I., Kim, S.H., Kim, J.H., Park, S.R., 2019. Composition and structure of the marine benthic community in Terra Nova Bay, Antarctica: Responses of the benthic assemblage to disturbances. PLoS One. https://doi.org/10.1371/journal.pone.0225551

Kans, J., 2020. Entrez direct: E-utilities on the UNIX command line, in: Entrez Programming Utilities Help [Internet]. National Center for Biotechnology Information (US).

Kanta, M.R., Babula, J., Anilkumar, N.P., Krishna, N.R., Bhaskar, P.V., Soares, M.A., 2017. Variability of chlorophyll-a and diatoms in the frontal ecosystem of Indian Ocean sector of the Southern Ocean. Polish Polar Res. 375–392. https://doi.org/10.1515/popore-2017-0014

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kelley, D., Richards, C., 2020. oce: Analysis of Oceanographic Data.

Kim, S., Hammerstrom, K., Dayton, P., 2019. Epifaunal community response to iceberg-mediated environmental change in mcmurdo sound, antarctica. Mar. Ecol. Prog. Ser. https://doi.org/10.3354/meps12899

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation

sequencing-based diversity studies. Nucleic Acids Res. 41, e1–e1. https://doi.org/10.1093/nar/gks808

Krzeminska, M., Kuklinski, P., 2018. Biodiversity patterns of rock encrusting fauna from the shallow sublittoral of the Admiralty Bay. Mar. Environ. Res. https://doi.org/10.1016/j.marenvres.2018.03.016

Kuklinski, P., Balazy, P., Krzemińska, M., Bielecka, L., 2017. Species pool structure explains patterns of Antarctic rock-encrusting organism recruitment. Polar Biol. https://doi.org/10.1007/s00300-017-2159-3

Lacoursière Roussel, A., Howland, K., Normandeau, E., Grey, E.K., Archambault, P., Deiner, K., Lodge, D.M., Hernandez, C., Leduc, N., Bernatchez, L., 2018. eDNA metabarcoding as a new surveillance approach for coastal Arctic biodiversity. Ecol. Evol. 8, 7763–7777. https://doi.org/10.1002/ece3.4213

Layton, K., 2012. Examining patterns of genetic variation in Canadian marine molluscs through DNA barcodes (Doctoral dissertation). http://hdl.handle.net/10214/5505

Layton, K. K., Corstorphine, E. A., & Hebert, P. D., 2016. Exploring Canadian echinoderm diversity through DNA barcodes. PloS one, 11(11), e0166118. https://doi.org/10.1371/journal.pone.0166118

Leray, M., Knowlton, N., 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1424997112

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., Boehm, J.T., Machida, R.J., 2013. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front. Zool. 10, 1–14.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, Q., Zhao, Q., McMinn, A., Yang, E.J., Jiang, Y., 2020. Planktonic microbial eukaryotes in polar surface waters: recent advances in high-throughput sequencing. Mar. Life Sci. Technol. 1–9. https://doi.org/10.1007/s42995-020-00062-y

Lo Giudice, A. Lo, Caruso, C., Mangano, S., Bruni, V., De Domenico, M., Michaud, L., 2012. Marine bacterioplankton diversity and community composition in an Antarctic coastal environment. Microb. Ecol. 63, 210–223. https://doi.org/10.1007/978-3-030-02786-5_6

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Lynn, D., 2008. The ciliated protozoa: characterization, classification, and guide to the literature. Springer Science & Business Media.

Machida, R.J., Leray, M., Ho, S.-L., Knowlton, N., 2017. Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. Sci. data 4, 1–7.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2015. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ 3, e1420.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., Dunthorn, M., 2014. Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2, e593.

Maldonado, M., Durfort, M., McCarthy, D. A., & Young, C. M. 2003. The cellular basis of photobehavior in the tufted parenchymella larva of demosponges. Marine Biology, 143(3), 427-441.

Mangoni, O., Saggiomo, M., Modigh, M., Catalano, G., Zingone, A., Saggiomo, V., 2009. The role of platelet ice microalgae in seeding phytoplankton blooms in Terra Nova Bay (Ross Sea,

Antarctica): a mesocosm experiment. Polar Biol. 32, 311–323. https://doi.org/10.1007/s00300-008-0507-z

Martin-Platero, A.M., Cleary, B., Kauffman, K., Preheim, S.P., McGillicuddy, D.J., Alm, E.J., Polz, M.F., 2018. High resolution time series reveals cohesive but short-lived communities in coastal plankton. Nat. Commun. https://doi.org/10.1038/s41467-017-02571-4

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 10–12. https://doi.org/10.14806/ej.17.1.200

Matsuoka, K., Skoglund, A., Roth, G., 2018. Quantarctica, Norwegian Polar Institute.

Meyer, C. P., 2003. Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biological Journal of the Linnean Society, 79(3), 401-459. https://doi.org/10.1046/j.1095-8312.2003.00197.x

McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Moline, M.A., Prezelin, B.B., 1996. Long-term monitoring and analyses of physical factors regulating variability in coastal Antarctic phytoplankton biomass, in situ productivity and taxonomic composition over subseasonal, seasonal and interannual time scales. Mar. Ecol. Prog. Ser. 145, 143–160. https://doi.org/10.3354/meps145143

Monti, M., Zoccarato, L., Umani, S.F., 2017. Microzooplankton composition under the sea ice and in the open waters in Terra Nova Bay (Antarctica). Polar Biol. 40, 891–901. https://doi.org/10.1007/s00300-016-2016-9

Moreira, D., López García, P., 2019. Time series are critical to understand microbial plankton diversity and ecology. Mol. Ecol. 28, 920–922. https://doi.org/10.1111/mec.15015

Munch, K., Boomsma, W., Huelsenbeck, J.P., Willerslev, E., Nielsen, R., 2008. Statistical assignment of DNA sequences using Bayesian phylogenetics. Syst. Biol. https://doi.org/10.1080/10635150802422316

Navarro, L.M., Fernández, N., Guerra, C., Guralnick, R., Kissling, W.D., Londoño, M.C., Muller-Karger, F., Turak, E., Balvanera, P., Costello, M.J., 2017. Monitoring biodiversity change through effective global coordination. Curr. Opin. Environ. Sustain. 29, 158–169. https://doi.org/10.1016/j.cosust.2018.02.005

Nieblas, A.-E., Sloyan, B.M., Hobday, A.J., Coleman, R., Richardsone, A.J., 2009. Variability of biological production in low wind forced regional upwelling systems: A case study off southeastern Australia. Limnol. Oceanogr. 54, 1548–1558. https://doi.org/10.4319/lo.2009.54.5.1548

Nilsson, R.H., Larsson, K.H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: Handling dark taxa and parallel taxonomic classifications. Nucleic Acids Res. https://doi.org/10.1093/nar/gky1022

Nuccio, C., Innamorati, M., Lazzara, L., Mori, G., Massi, L., 2000. Spatial and temporal distribution of phytoplankton assemblages in the Ross Sea, in: Ross Sea Ecology. Springer, pp. 231–245. https://doi.org/10.1007/978-3-642-59607-0_19

Obst, M., Exter, K., Allcock, A.L., Arvanitidis, C., Axberg, A., Bustamante, M., Cancio, I., Carreira-Flores, D., Chatzinikolaou, E., Chatzigeorgiou, G., Chrismas, N., Clark, M.S., Comtet, T., Dailianis, T., Davies, N., Deneudt, K., de Cerio, O.D., Fortič, A., Gerovasileiou, V., Hablützel, P.I., Keklikoglou, K., Kotoulas, G., Lasota, R., Leite, B.R., Loisel, S., Lévêque, L., Levy, L., Malachowicz, M., Mavrič, B., Meyer, C., Mortelmans, J., Norkko, J., Pade, N., Power, A.M., Ramšak, A., Reiss, H., Solbakken, J., Staehr, P.A., Sundberg, P., Thyrring, J., Troncoso, J.S., Viard, F., Wenne, R., Yperifanou, E.I., Zbawicka, M., Pavloudi, C., 2020. A Marine Biodiversity

Observation Network for Genetic Monitoring of Hard-Bottom Communities (ARMS-MBON). Front. Mar. Sci. https://doi.org/10.3389/fmars.2020.572680

Ocean Biology Processing Group, 2018a. Moderate-Resolution Imaging Spectroradiometer (MODIS) Aqua Chlorophyll Data Reprocessing.

Ocean Biology Processing Group, 2018b. Moderate-Resolution Imaging Spectroradiometer (MODIS) Aqua Particulate Organic Carbon Data Reprocessing.

Pabi, S., Arrigo, K.R., 2006. Satellite estimation of marine particulate organic carbon in waters dominated by different phytoplankton taxa. J. Geophys. Res. Ocean. 111. https://doi.org/10.1029/2005JC003137

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. Bioinformatics 35, 526–528.

Park, J., Kim, J., Kim, H., Hwang, J., Jo, Y., Lee, S.H., 2019. Environmental forcings on the remotely sensed phytoplankton bloom phenology in the central Ross Sea Polynya. J. Geophys. Res. Ocean. 124, 5400–5417. https://doi.org/10.1029/2019JC015222

Pawlowski, J., Apothéloz-Perret-Gentil, L., Altermatt, F., 2020. Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. Mol. Ecol. https://doi.org/10.1111/mec.15643

Pearman, J.K., Anlauf, H., Irigoien, X., Carvalho, S., 2016. Please mind the gap - Visual census and cryptic biodiversity assessment at central Red Sea coral reefs. Mar. Environ. Res. https://doi.org/10.1016/j.marenvres.2016.04.011

Pearman, J.K., Chust, G., Aylagas, E., Villarino, E., Watson, J.R., Chenuil, A., Borja, A., Cahill, A.E., Carugati, L., Danovaro, R., David, R., Irigoien, X., Mendibil, I., Moncheva, S., Rodríguez-Ezpeleta, N., Uyarra, M.C., Carvalho, S., 2020. Pan-regional marine benthic cryptobiome biodiversity patterns revealed by metabarcoding Autonomous Reef Monitoring Structures. Mol. Ecol. https://doi.org/10.1111/mec.15692

Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. R J. 10, 439–446. https://doi.org/10.32614/RJ-2018-009

Peck, L.S., 2018. ANTARCTIC MARINE BIODIVERSITY: ADAPTATIONS, ENVIRONMENTS AND RESPONSES TO CHANGE.

Penev, L., Mietchen, D., Chavan, V., Hagedorn, G., Smith, V., Shotton, D., Ó Tuama, É., Senderov, V., Georgiev, T., Stoev, P., Groom, Q., Remsen, D., Edmunds, S., 2017. Strategies and guidelines for scholarly publishing of biodiversity data. Res. Ideas Outcomes. https://doi.org/10.3897/rio.3.e12431

Pentinsaari, M., Ratnasingham, S., Miller, S.E., Hebert, P.D.N., 2020. BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries? PLoS One. https://doi.org/10.1371/journal.pone.0231814

Piazza, P., Błażewicz-Paszkowycz, M., Ghiglione, C., Alvaro, M.C., Schnabel, K., Schiaparelli, S., 2014. Distributional records of Ross Sea (Antarctica) tanaidacea from museum samples stored in the collections of the Italian National Antarctic Museum (MNA) and the New Zealand National Institute of Water and Atmospheric Research (NIWA). Zookeys. https://doi.org/10.3897/zookeys.451.8373

Piñol, J., Senar, M.A., Symondson, W.O.C., 2019. The choice of universal primers and the characteristics of the species mixture determine when DNA metabarcoding can be quantitative. Mol. Ecol. 28, 407–419. https://doi.org/10.1111/mec.14776

Pope, L.C., Liggins, L., Keyse, J., Carvalho, S.B., Riginos, C., 2015. Not the time or the place: The missing spatio-temporal link in publicly available genetic data. Mol. Ecol. https://doi.org/10.1111/mec.13254

Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R.D., 2017. Global biodiversity monitoring: from data sources

to essential biodiversity variables. Biol. Conserv. 213, 256–263. https://doi.org/10.1016/j.biocon.2016.07.014

QGIS Development Team, 2020. QGIS Geographic Information System. Open Source Geospatial Foundation Project.

Qu, B., Gabric, A.J., Zeng, M., Liu, X., 2020. Correlations among phytoplankton biomass, sea ice and wind speed in Barents Sea and the future climate trends. Polar Sci. 100525. https://doi.org/10.1016/j.polar.2020.100525

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596. https://doi.org/10.1093/nar/gks1219

Questel, J.M., Hopcroft, R.R., DeHart, H.M., Smoot, C.A., Kosobokova, K.N., Bucklin, A., 2021. Metabarcoding of zooplankton diversity within the Chukchi Borderland, Arctic Ocean: improved resolution from multi-gene markers and region-specific DNA databases. Mar. Biodivers. https://doi.org/10.1007/s12526-020-01136-x

R Core Team, 2020. R: A Language and Environment for Statistical Computing.

Rajaram, S., Oono, Y., 2010. NeatMap-non-clustering heat map alternatives in R. BMC Bioinformatics 11, 1–9. https://doi.org/10.1186/1471-2105-11-45

Ransome, E., Geller, J.B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A.G., Meyer, C.P., 2017. The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. PLoS One 12, e0175066. https://doi.org/10.1371/journal.pone.0175066

Ratnasingham, S., Hebert, P.D.N., 2007. Barcoding. BOLD barcode life data Syst. (www. barcodinglife. org). Mol. Ecol. Notes 7, 355–364.

Riesgo, A., Taboada, S., Avila, C., 2015. Evolutionary patterns in Antarctic marine invertebrates: An update on molecular studies. Mar. Genomics. https://doi.org/10.1016/j.margen.2015.07.005

Rodriguez-Ezpeleta, N., Morissette, O., Bean, C., Manu, S., Banerjee, P., Lacoursiere, A., Ben, K., Alter, E., Roger, F., Holman, L., 2020. Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on “Environmental DNA: What’s behind the term?” by Pawlowski et al.(2020). Authorea Prepr.

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584

Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P.M., Bourgoin, T., DeWalt, R.E., Decock, W., van Nieukerken, E.J., Penev, L., 2020. Species 2000 & ITIS Catalogue of Life, 2020-12-01. Species 2000 Nat. Leiden, Netherlands.

Rosso, A., Sanfilippo, R., 2000. Shallow-Water Bryozoans and Serpuloideans from the Ross Sea (Terra Nova Bay, Antarctica), in: Ross Sea Ecology. https://doi.org/10.1007/978-3-642-59607-0_37

Russel, J., 2021. MicEco: Various functions for microbial community data.

Saggiomo, M., Poulin, M., Mangoni, O., Lazzara, L., De Stefano, M., Sarno, D., Zingone, A., 2017. Spring-time dynamics of diatom communities in landfast and underlying platelet ice in Terra Nova Bay, Ross Sea, Antarctica. J. Mar. Syst. 166, 26–36. https://doi.org/10.1016/j.jmarsys.2016.06.007

Sallée, J.-B., Llort, J., Tagliabue, A., Lévy, M., 2015. Characterization of distinct bloom phenology regimes in the Southern Ocean. ICES J. Mar. Sci. 72, 1985–1998. https://doi.org/10.1093/icesjms/fsv069

Sazhin, A.F., Artigas, L.F., Nejstgaard, J.C., Frischer, M.E., 2007. The colonization of two Phaeocystis species (Prymnesiophyceae) by pennate diatoms and other protists: a significant

contribution to colony biomass, in: Phaeocystis, Major Link in the Biogeochemical Cycling of Climate-Relevant Elements. Springer, pp. 137–145. https://doi.org/10.1007/s10533-007-9086-2

Schiaparelli, S., Alvaro, M.C., Cecchetto, M., Guzzi, A., Piazza, P., 2018. Dalla rilevanza nazionale a quella internazionale: le strategie adottate dal Museo Nazionale dell'Antartide (MNA, Sede di Genova).

Schiaparelli, S., Danis, B., Wadley, V., Michael Stoddart, D., 2013. The Census of Antarctic Marine Life: The First Available Baseline for Antarctic Marine Biodiversity. https://doi.org/10.1007/978-3-642-27349-0_1

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Scott Chamberlain, Eduard Szocs, 2013. taxize - taxonomic search and retrieval in R. F1000Research.

Secretariat, G., 2019. GBIF backbone taxonomy. Checkl. Dataset [cited 2021 Feb 7]. https://doi.org/10.15468/39omei

Sewell, M., Lavery, S., Baker, C., 2006. Whose larva is that? Molecular identification of planktonic larvae of the Ross Sea. New Zeal. Aquat. Environ. Biodivers. Rep.

Sewell, M.A., Jury, J.A., 2011. Seasonal patterns in diversity and abundance of the High Antarctic meroplankton: plankton sampling using a Ross Sea desalination plant. Limnol. Oceanogr. 56, 1667–1681. https://doi.org/10.4319/lo.2011.56.5.1667

Sewell, M.A., Jury, J.A., 2009. Desalination plants as plankton sampling devices in temporal studies: proof of concept and suggestions for the future. Limnol. Oceanogr. Methods 7, 363–370. https://doi.org/10.4319/lom.2009.7.363

Shearer, T.L., Van Oppen, M.J.H., Romano, S.L., Wörheide, G., 2002. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol. Ecol. https://doi.org/10.1046/j.1365-294X.2002.01652.x

Sibbett, B., Rieseberg, L.H., Narum, S., 2020. The Genomic Observatories Metadatabase. Mol. Ecol. Resour. https://doi.org/10.1111/1755-0998.13283

signal developers, 2014. {signal}: Signal processing.

Sire, L., Gey, D., Debruyne, R., Noblecourt, T., Soldati, F., Barnouin, T., Parmain, G., Bouget, C., Lopez-Vaamonde, C., Rougerie, R., 2019. The Challenge of DNA Barcoding Saproxylic Beetles in Natural History Collections—Exploring the Potential of Parallel Multiplex Sequencing With Illumina MiSeq. Front. Ecol. Evol. https://doi.org/10.3389/fevo.2019.00495

Smith Jr, W.O., Dennett, M.R., Mathot, S., Caron, D.A., 2003. The temporal dynamics of the flagellated and colonial stages of Phaeocystis antarctica in the Ross Sea. Deep Sea Res. Part II Top. Stud. Oceanogr. 50, 605–617. https://doi.org/10.1016/S0967-0645(02)00586-6

Smith, W.O., Ainley, D.G., Cattaneo-Vietti, R., 2007. Trophic interactions within the Ross Sea continental shelf ecosystem. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2006.1956

South, A., 2017. rnaturalearth: World Map Data from Natural Earth.

Stanwell-Smith, D., Barnes, D.K.A., 1997. Benthic community development in Antarctica: Recruitment and growth on settlement panels at Signy Island. J. Exp. Mar. Bio. Ecol. 212, 61–79. https://doi.org/10.1016/S0022-0981(96)02754-2

Stark, J.S., 2008. Patterns of higher taxon colonisation and development in sessile marine benthic assemblages at Casey Station, Antarctica, and their use in environmental monitoring. Mar. Ecol. Prog. Ser. 365, 77–89. https://doi.org/10.3354/meps07559

Stein, E. D., White, B. P., Mazor, R. D., Miller, P. E., & Pilgrim, E. M., 2013. Evaluating ethanol-based sample preservation to facilitate use of DNA barcoding in routine freshwater

biomonitoring programs using benthic macroinvertebrates. PloS one, 8(1), e51273. • https://doi.org/10.1371/journal.pone.0051273

Straube, D., Juen, A., 2013. Storage and shipping of tissue samples for DNA analyses: A case study on earthworms. Eur. J. Soil Biol. 57, 13–18. https://doi.org/10.1016/j.ejsobi.2013.04.001

Suter, L., Polanowski, A.M., Clarke, L.J., Kitchener, J.A., Deagle, B.E., 2020. Capturing open ocean biodiversity: Comparing environmental DNA metabarcoding to the continuous plankton recorder. Mol. Ecol. https://doi.org/10.1111/mec.15587

Taberlet, P., Bonin, A., Coissac, E., Zinger, L., 2018. Environmental DNA: For biodiversity research and monitoring. Oxford University Press. https://doi.org/10.1093/oso/9780198767220.001.0001

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., Willerslev, E., 2012. Towards next generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. 21, 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Thatje, S., Hillenbrand, C.D., Larter, R., 2005. On the origin of Antarctic marine benthic community structure. Trends Ecol. Evol. https://doi.org/10.1016/j.tree.2005.07.010

Thrush, S., Dayton, P., Cattaneo-Vietti, R., Chiantore, M., Cummings, V., Andrew, N., Hawes, I., Kim, S., Kvitek, R., Schwarz, A.M., 2006. Broad-scale factors influencing the biodiversity of coastal benthic communities of the Ross Sea. Deep. Res. Part II Top. Stud. Oceanogr. https://doi.org/10.1016/j.dsr2.2006.02.006

Torstensson, A., Dinasquet, J., Chierici, M., Fransson, A., Riemann, L., Wulff, A., 2015. Physicochemical control of bacterial and protist community composition and diversity in A ntarctic sea ice. Environ. Microbiol. 17, 3869–3881. https://doi.org/10.1111/1462-2920.12865

Tripathy, S.C., Jena, B., 2019. Iron-Stimulated Phytoplankton Blooms in the Southern Ocean: a Brief Review. Remote Sens. Earth Syst. Sci. 2, 64–77. https://doi.org/10.1007/s41976-019-00012-y

Troudet, J., Vignes-Lebbe, R., Grandcolas, P., Legendre, F., 2018. The increasing disconnection of primary biodiversity data from specimens: How does it happen and how to handle it? Syst. Biol. https://doi.org/10.1093/sysbio/syy044

Truzzi, C., Annibaldi, A., Finale, C., Libani, G., Romagnoli, T., Scarponi, G., Illuminati, S., 2015. Separation of micro-phytoplankton from inorganic particulate in Antarctic seawater (Ross Sea) for the determination of Cd, Pb and Cu: optimization of the analytical methodology. Anal. Methods 7, 5490–5496. https://doi.org/10.1039/C5AY00730E

Turon, X., Antich, A., Palacín, C., Præbel, K., Wangensteen, O.S., 2020. From metabarcoding to metaphylogeography: separating the wheat from the chaff. Ecol. Appl. https://doi.org/10.1002/eap.2036

Vacchi, M., DeVries, A.L., Evans, C.W., Bottaro, M., Ghigliotti, L., Cutroneo, L., Pisano, E., 2012. A nursery area for the Antarctic silverfish Pleuragramma antarcticum at Terra Nova Bay (Ross Sea): First estimate of distribution and abundance of eggs and larvae under the seasonal sea-ice. Polar Biol. https://doi.org/10.1007/s00300-012-1199-y

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P.F., Bellemain, E., Besnard, A., Coissac, E., Boyer, F., 2016. Next generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol. Ecol. 25, 929–942. https://doi.org/10.1111/mec.13428

van Soest, R.W.M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., de Voogd, N.J., Santodomingo, N., Vanhoorne, B., Kelly, M., Hooper, J.N.A., 2012. Global diversity of sponges (Porifera). PLoS One. https://doi.org/10.1371/journal.pone.0035105

Vargas, S., Kelly, M., Schnabel, K., Mills, S., Bowden, D., Wörheide, G., 2015. Diversity in a cold hot-spot: DNA-barcoding reveals patterns of evolution among Antarctic demosponges (class demospongiae, phylum Porifera). PLoS One. https://doi.org/10.1371/journal.pone.0127573

Vasselon, V., Domaizon, I., Rimet, F., Kahlert, M., Bouchez, A., 2017. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshw. Sci. 36, 162–177. https://doi.org/10.1086/690649

Veerapaneni, S., Long, B., Freeman, S., Bond, R., 2007. Reducing energy consumption for seawater desalination. Journal American Water Work. Assoc. 99, 95–106. https://doi.org/10.1002/j.1551-8833.2007.tb07958.x

Villacorte, L.O., Tabatabai, S.A.A., Anderson, D.M., Amy, G.L., Schippers, J.C., Kennedy, M.D., 2015. Seawater reverse osmosis desalination and (harmful) algal blooms. Desalination 360, 61–80. https://doi.org/10.1016/j.desal.2015.01.007

Wang, E., Cook, D., Hyndman, R.J., 2020. A new tidy data structure to support exploration and modeling of temporal data. J. Comput. Graph. Stat. 1–13. https://doi.org/10.1080/10618600.2019.1695624

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73, 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wangensteen, O.S., Palacín, C., Guardiola, M., Turon, X., 2018. DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. PeerJ 6, e4705.

Ward, R. D., Holmes, B. H., & O'HARA, T. D., 2008. DNA barcoding discriminates echinoderm species. Molecular Ecology Resources, 8(6), 1202-1211. https://doi.org/10.1111/j.1755-0998.2008.02332.x

Weigand, H., Beermann, A.J., Čiampor, F., Costa, F.O., Csabai, Z., Duarte, S., Geiger, M.F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A.M., Willassen, E., Wyler, S.A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., Dijkstra, K.D., Eisendle, U., Freyhof, J., Gadawski, P., Graf, W., Haegerbaeumer, A., van der Hoorn, B.B.,

Japoshvili, B., Keresztes, L., Keskin, E., Leese, F., Macher, J., Mamos, T., Paz, G., Pešić, V., Pfannkuchen, D.M., Pfannkuchen, M.A., Price, B.W., Rinkevich, B., Teixeira, M.A.L., Várbíró, G., Ekrem, T., 2019. DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. bioRxiv. https://doi.org/10.1101/576553

West, K.M., Stat, M., Harvey, E.S., Skepper, C.L., DiBattista, J.D., Richards, Z.T., Travers, M.J., Newman, S.J., Bunce, M., 2020. eDNA metabarcoding survey reveals fine scale coral reef community variation across a remote, tropical island ecosystem. Mol. Ecol. 29, 1069–1086. https://doi.org/10.1111/mec.15382

Westfall, K.M., Therriault, T.W., Abbott, C.L., 2020. A new approach to molecular biosurveillance of invasive species using DNA metabarcoding. Glob. Chang. Biol. https://doi.org/10.1111/gcb.14886

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., Fran<U+00E7>ois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., M<U+00FC>ller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019a. Welcome to the {tidyverse}. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Francois, R., Henry, L., Müller, K., 2019b. Package dplyr: A Grammar of Data Manipulation. R package version 0.8. 5.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., Vieglais, D., 2012. Darwin core: An evolving community-developed biodiversity data standard. PLoS One. https://doi.org/10.1371/journal.pone.0029715

Wilke, C.O., 2020. cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”

Wilkins, D., Yau, S., Williams, T.J., Allen, M.A., Brown, M. V, DeMaere, M.Z., Lauro, F.M., Cavicchioli, R., 2013. Key microbial drivers in Antarctic aquatic environments. FEMS Microbiol. Rev. 37, 303–335. https://doi.org/10.1111/1574-6976.12007

Wolf, P.H., Siverns, S., Monti, S., 2005. UF membranes for RO desalination pretreatment. Desalination 182, 293–300. https://doi.org/10.1016/j.desal.2005.05.006

Wright, E.S., 2016. Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. R J. 8, 352–359.

Zaiko, A., Schimanski, K., Pochon, X., Hopkins, G.A., Goldstien, S., Floerl, O., Wood, S.A., 2016. Metabarcoding improves detection of eukaryotes from early biofouling communities: implications for pest monitoring and pathway management. Biofouling. https://doi.org/10.1080/08927014.2016.1186165

Zhang, G.K., Chain, F.J.J., Abbott, C.L., Cristescu, M.E., 2018. Metabarcoding using multiplexed markers increases species detection in complex zooplankton communities. Evol. Appl. 11, 1901–1914. https://doi.org/10.1111/eva.12694

Zimmerman, T.L., Martin, J.W., 2004. Artificial Reef Matrix Structures (Arms): An Inexpensive and Effective Method for Collecting Coral Reef-Associated Invertebrates. Gulf Caribb. Res. https://doi.org/10.18785/gcr.1601.08

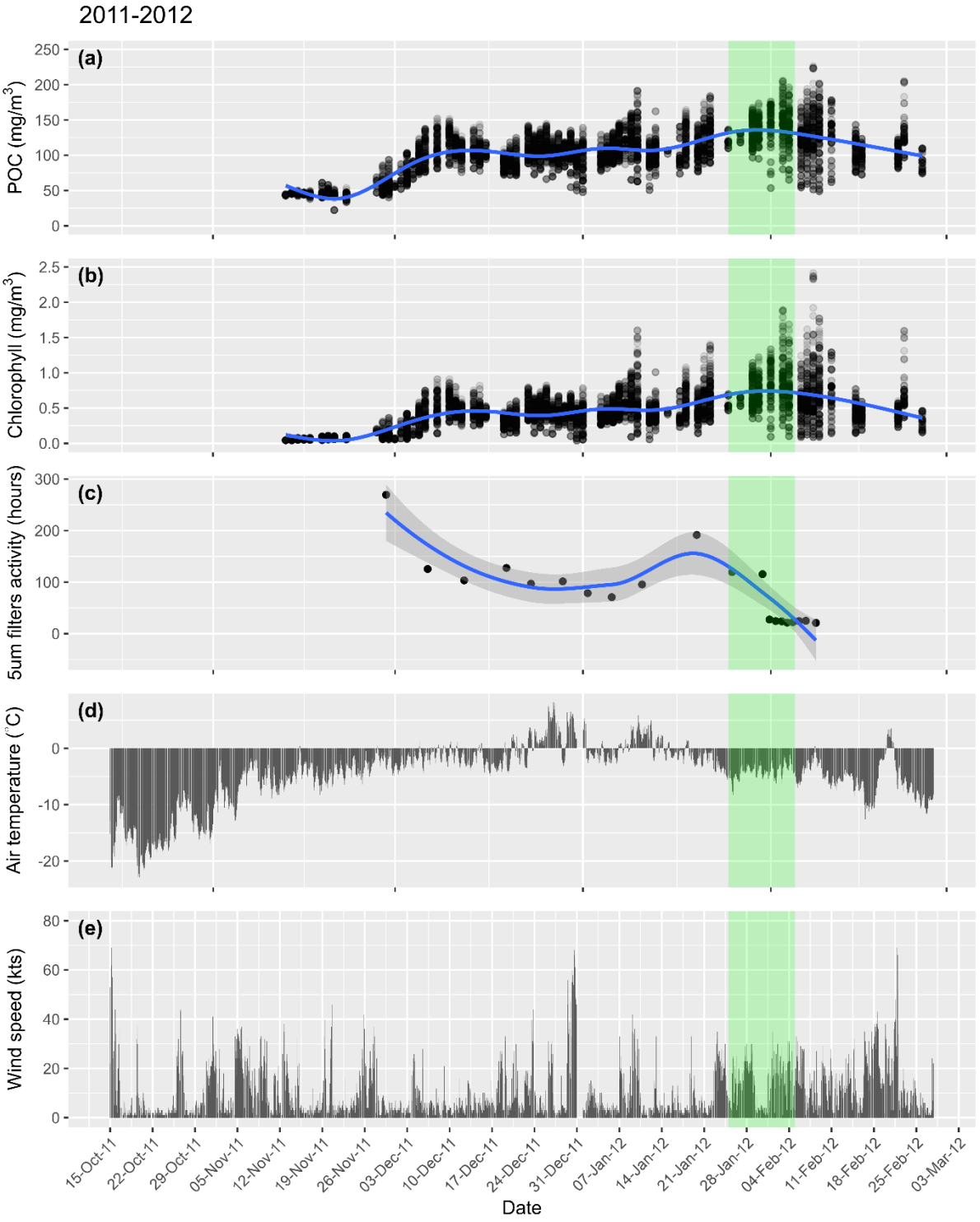
Zizka, A., Carvalho, F.A., Calvente, A., Baez-Lizarazo, M.R., Cabral, A., Ramos Coelho, J.F., Colli-Silva, M., Fantinati, M.R., Fernandes, M.F., Ferreira-Araújo, T., Lambert Moreira, F.G., da Cunha Santos, N.M., Borges Santos, T.A., dos Santos-Costa, R.C., Serrano, F.C., da Silva, A.P.A., de Souza Soares, A., de Souza, P.G.C., Tomaz, E.C., Vale, V.F., Vieira, T.L., Antonelli, A., 2020. No one-size-fits-all solution to clean GBIF. PeerJ. https://doi.org/10.7717/peerj.9916

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V., Antonelli, A., 2019.

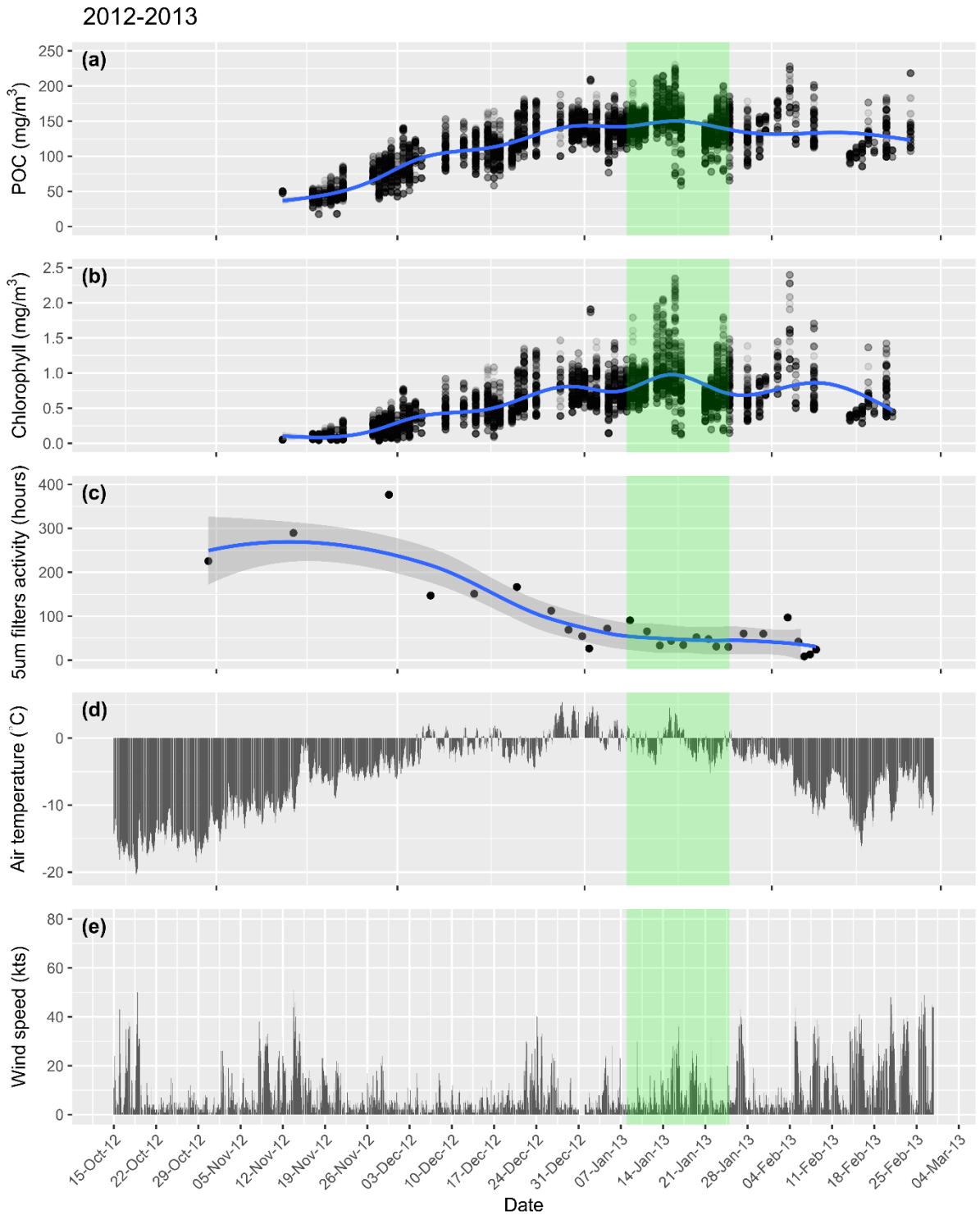
CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.13152

Zoccarato, L., Pallavicini, A., Cerino, F., Umani, S.F., Celussi, M., 2016. Water mass dynamics shape Ross Sea protist communities in mesopelagic and bathypelagic layers. Prog. Oceanogr. 149, 16–26. https://doi.org/10.1016/j.pocean.2016.10.003

Appendix figures:



Appendix Figure 1: Daily satellite recordings of (a) Particulate Organic Carbon (POC) and (b) Chlorophyll, measured in milligrams per cubic meter. (c) 5 µm cartridge filters activity time measured in hours. Hourly recordings of AWS "Eneide" on (d) surface temperature and (e) wind speed for the MZS opening season 2011-2012. Blue lines are the smoothing lines, with the grey areas depicting the confidence intervals around the smoothing, according to a generalized additive model for (a) and (b), and to a loess approximation for (c). Green shaded areas highlight the investigated time range, from the installation day of the first filter to the sampling day of the last one.



Appendix Figure 2: Daily satellite recordings of (a) Particulate Organic Carbon (POC) and (b) Chlorophyll, measured in milligrams per cubic meter. (c) 5 µm cartridge filters activity time measured in hours. Hourly recordings of AWS "Eneide" on (d) surface temperature and (e) wind speed for the MZS opening season 2012-2013. Blue lines are the smoothing lines, with the grey areas depicting the confidence intervals around the smoothing, according to a generalized additive model for (a) and (b), and to a loess approximation for (c). Green shaded areas highlight the investigated time range, from the installation day of the first filter to the sampling day of the last one.